Semantic structures, communicative strategies and the emergence of language

Semantic Structures, Communicative Strategies and the Emergence of Language

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: *silent baboon* by Marieke Schouwstra

ISBN: 978-94-6093-096-6
NUR: 616

Semantic Structures, Communicative Strategies and the Emergence of Language

De rol van semantiek en communicatie in het ontstaan van taal

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 30 november 2012 des middags te 2.30 uur

door

Marieke Schouwstra

geboren 2 april 1981
te Harderwijk

Promotor: Prof. dr. H.E. de Swart

Contents

Acknowledgements

There are many ways to thank people, and I wish I had written down the most elegant ones when I was still in my right mind. Because now, here I am, not long before the deadline searching for words... Well, what the heck.

First of all, I would like to thank my supervisor, Henriëtte de Swart. Henriëtte, when you suggested that I could do a project on language evolution, I wasn't immediately enthusiastic. In fact, I spent an afternoon shaking my head, thinking: 'I'm not going to do that evolutionary stuff'. And look at me now! I really, really enjoyed working on this topic, and I really appreciate how you encouraged me to follow my instincts, and helped me to organise my ideas and put them into words. And thank you very much for your support and your positive attitude when the road got bumpy.

I am very grateful to the members of my reading committee: Simon Kirby, Rudie Botha, Frank Veltman, Martin Everaert, and Herman Hendriks. Thank you all for reading my dissertation and commenting on it.

The Utrecht institute of Linguistics OTS is a very lively and friendly environment, and I am very happy to have done my PhD there. Even though I was the only PhD student working on language evolution, there were many colleagues I could talk to and discuss issues with. I would like to name a few of them. Jakub Dotlacil, thanks for being so patient to explain (sometimes even very basic) linguistics stuff to me. Rick Nouwen, thanks for advice on all kinds of things; sorry for bursting into your office so many times to ask things! Marijn Struiksma, thank you for explaining statistics properly. Arnout Koornneef, thank you for your help with the setup of my last experiment. (And I never told you, but your dissertation is written in such a contagiously enthusiastic way, that I read passages from it every now an then to remind me that writing up can be really fun.) Alexis Dimitriadis, thank you for providing the LaTeXclass for my dissertation. Other people who have helped me increase my understanding of the various subdisciplines of linguistics (and a bunch of areas outside linguistics) include Joost Zwarts, Boban Arsenijevic, Gianluca Giorgolo, Bert Le Bruyn, Tom Lentz, Andreas Pankau.

Further, I would like to thank all my UiL OTS office mates (ordered chronologically): Joke de Lange, Jakub Dotlacil, Natalia Slioussar, Gaetano Fiorin, Marijana Marelj, Anna Chernilovskaya, Ora Matushansky, Xiaoli Dong, Marko Simonovic, Ana Aguilar Guevara, and Carolien van den Hazelkamp. Thanks for your support, for orange and coffee breaks together, and for tolerating my afternoon power naps on the sofa when I was pregnant.

Describing the friendly environment at the UiL OTS, I should also mention the lunches, Friday drinks and other pub sessions, table football games, cups of coffee and nice chats in the corridor. I am a bit of a wanderer, and the list below is probably incomplete... Please know that I did appreciate all the good moments! Thank you: Min Que, Arjen Zondervan, Øystein Nilsen, Clizia Welker, Sander van der Harst, Hannah de Mulder, Christina Unger, Roberta Tedeschi, Natalie Boll-Avetisyan, Naomi Kamoen, Anna Volkova, Frans Adriaans, Radek Simik, Nadya Goldberg, Lizet van Ewijk, Loes Koring, Bettina Gruber, Desiree Capel, Marko Hladnik, Marijke de Belder, Rianne Schippers, Liv Persson, Sophia Katrenko, Brigitta Keij, Sandrien van Ommen, Kiki Kushartanti, Mulusew Wondem.

Finally, thank you to Maaike Schoorlemmer, and the UiL OTS secretaries for helping me when necessary.

Outside the institute, there were many people who helped me. Ruth Millikan, thank you for discussing your work with me and for influencing my thinking about thought. I hope you don't disagree too much with the way in which I have isolated and described a small fraction of your work. Thank you to Jelle Zuidema and Bart de Boer for introducing me to many new people at my first Evolang conference and for making me feel more at home in evolutionary linguistics. Susan Goldin-Meadow and Ray Jackendoff, thank you for taking the time to comment on my work and answer my questions. Tikitu de Jager, thank you very much for proofreading my thesis, and for writing such funny comments that even the last (and most tired) stages of editing were fun. Henk Zeevat, thanks for persuading me to come along to Szklarska at a time (long ago) when I had hardly an idea of what academia had to offer.

During my project I visited the Language Evolution and Computation research unit at the University of Edinburgh for a semester, and I am very happy to have stayed there. Not only did I meet many, many interesting people, I also started doing lab experiments there. I would like to thank Jim Hurford for being my academic sponsor during this visit, and for taking the time to discuss many issues. Furthermore, I am very much indebted to Simon Kirby for encouraging me to do experiments and for his tremendous help, inspiration and encouraging words, also after my Edingburgh visit. Thanks! And thanks very much to Mónica Tamariz for help, advice and a couple of nudges in the right direction at crucial moments (for example, during the CogSci conference). To all other PhD students and postdocs I met in Edinburgh: thank you very much for discussing your work with me, for the Friday pints, the jaunts up Arthur's Seat and other hills, the babel meetings, and for making me feel very much at

home in Scotland.

I supervised a number of students, four of whom deserve special mention: Anouschka van Leeuwen, Marianne Smit, Nicky Marien and Kimberly Snoyl, thank you for joining my merry adventure in language evolution experiments. You all worked with enthusiasm and dedication, and I am grateful for the work you did.

Outside my academic life there were many people without whom I would never have finished this dissertation. First of all, I would like to thank all my friends for all the fun moments, cups of coffee, dinners and contemporary poetry that pulled me through hard times, and for being my friends, even though I was sometimes absent, or absent-minded. Carola en Liefke, bedankt dat jullie mijn paranimfen willen zijn!

Many thanks to all three of Gijsbert's grannies for babysitting and helping out in our messy household.

I would like to thank my parents Wietze and Tineke, and my sisters Sandra and Laura. Bedankt dat jullie er waren zoals jullie er altijd waren: als familie, en zonder voorwaarden. Ik weet dat jullie er zijn, en dat jullie trots op me zijn.

Finally, I would like to thank Niels, who has done more for me than I could ever thank him for, and Gijsbert, who joined us and made us proud. Mijn lieve mannen, ik vind het fantastisch met jullie.

CHAPTER 1

Meaning and protolanguage

1.1 Introduction: the role of meaning in the emergence of language

1.1.1 Language and structure

This dissertation is about the emergence of language. The most striking property of human language is probably its structural complexity. Have a look at the following examples:

(1) Which movie does Susan imagine that Sarah saw last night?

(2) Who does Sam know a girl who is in love with?

Sentence (1) is correct English, but sentence (2) is not. And the following three lines from the English nursery rhyme 'The house that Jack built' show that very long sentences can be built in a language like English:

(3) a. This is the house that Jack built.
b. This is the malt that lay in the house that Jack built.
c. This is the rat that ate the malt that lay in the house that Jack built.

Speakers of English will be able to understand not only sentence (3-a), but also (3-c). These examples only give us a hint of how structurally complex human language is. In linguistics, successful accounts have been put forward that explain the difference between (1) and (2), and that model how sentences

like (3-c) are produced or understood. The question why humans can do these things and animals cannot, however, is still largely unanswered.

It is hard to imagine how a system that is so sophisticated and formally complex could be the product of evolution. To date, many scholars have assumed that syntactic complexity as we see it in modern language emerged not gradually but in one big leap (see, e.g., Berwick (2011)). To me, this assumption is not satisfactory, because giant leaps are rare in evolution, and as long as we have no reasons to refute the gradual emergence of language, we should investigate its possibility. Moreover, postulating that language came into existence with the rapid emergence of syntactic complexity puts syntax in a rather special position and does not do justice to observations that the systems necessary for language are deeply intertwined with other cognitive systems (Fitch, 2010; Hurford, 2007; Jackendoff, 2002).

Thus, even though the question how language became so structurally complex is tantalising, we should keep in mind that language is a many-faceted phenomenon, and that focusing on its structural complexity alone cannot result in a full understanding of the emergence of language (Fitch, 2010, p. 3).

In this dissertation, I will look at language from a different angle: that of language as a means to *share thoughts and meanings*. This point of view does not address the structural complexity of language directly, but it will allow me to say something about it eventually (see chapter 7).

1.1.2 Language and meaning

In everyday situations, we use language to convey meanings, and these meanings can be quite complex. For example, when I am in my office, I can ask my colleague to open the window, but I can also tell her about my weekend. I can tell her about the things that happened and what I thought about them. About what would have happened if things had gone differently. Or about what I would like to do *next* weekend.

Silly as these examples may seem, they do show the different ways in which language can be sophisticated. Utterances usually refer not just to one object in the world, but to combinations of things. And they allow us to share information in an abstract way. To see this, let us look at the examples above in more detail. If I ask my colleague to open the window, this is something that is here and now: the window is closed, and I want it to be open. If for some reason I had to convey my message without making any sound, I could indicate what I wanted by pointing at the window and my colleague would probably still understand me. If, however, I talk about the things I did last weekend, this is a bit more abstract. The things I talk about are not present in the room. Moreover, the report of my activities can be quite complex, in the sense that I can refer to different events that all bear certain relations to each other. For example, if I say 'I went to the pub with a friend and we had a good chat about life and all that,' my colleague will understand that the chatting took place during the

pub visit and not after. In other words, I describe two events, the second of which took place *during* the first.

Next, imagine that I talk to my colleague about things that could have happened if certain things had gone differently from the way they did: 'If I hadn't forgotten my raincoat yesterday, I wouldn't have come home soaking wet.' This is, in a way, even more complex: I am talking about things that did not actually take place. Yet, my colleague understands what I am talking about and she will probably give me a pitying look. When I talk about my plans for next weekend ('I'm going to write all weekend'), my colleague will know what I mean as well, even though I am talking about something utterly abstract: an event in the future.

These apparently simple examples show that we use language to do quite sophisticated things: we talk about things and events in the world. These things and events can be here or somewhere else, they can take place now, in the past, or in the future (or in some world different from the actual world). In other words, humans can 'build' abstract meanings and communicate them, without necessarily having the topic of their conversation in their vicinity.

Humans are the only species that build utterances that are at the same time meaningful and complex. Other animals do communicate with each other, and their communication can be used to refer to things in the world (as in, e.g., the warning calls of cotton-top tamarins, where different calls refer to different predators) or to make complex utterances (as in the elaborate songs of songbirds, which do have sophisticated structure, but no —referential— meaning), but the combination of meaningfulness and complexity is uniquely human (Hurford, 2011).

Thus, we have observed that human language is syntactically complex, but it is semantically complex as well. In this dissertation, I will show that we can address the origins of syntactic complexity by focusing on semantic complexity. I will do this by investigating a potential intermediate stage in the emergence of language, which I will call *semantic protolanguage*.

1.1.3 Semantic protolanguage

The initial idea for a semantic account of protolanguage was put forward by Jackendoff (2002), who claimed that sentences in protolanguage were governed by semantic and pragmatic principles. In other words, according to Jackendoff, protolanguage speakers made multi-word utterances, and these were not organised by strict syntactic rules, but instead organised by principles that are semantic or pragmatic in nature. An example of a semantic organisational principle is *AgentFirst*: to put the element that has most control first in a sentence.[1]

Jackendoff's hypothesis that meaning played a role in the structuring of utterances in protolanguage is not mere speculation. Following Bickerton (1990)'s

[1]For a more extensive description, see below in section 1.4.

lead, he points out that evidence for such a stage can be found in modern linguistic phenomena like pidgin languages, homesign systems, or the language of adults learning a second language. These phenomena all emerge in situations where people cannot use or learn a language normally. The organising principles that are observed are independent from their surrounding languages and are said to mirror principles from protolanguage (Jackendoff, 2002).

If there has indeed been a semantic stage in the emergence of language, the principles by which utterances are organised in such a stage can be seen as precursors of syntactic rules. In other words, syntactic rules have emerged out of semantic organising principles. In this dissertation I will investigate the semantic protolanguage hypothesis. I will discuss and evaluate the evidence that has been put forward in favour of it, and address the recommendation put forward by Botha (2003), that a non-ad hoc bridge theory should be formulated that takes us from modern linguistic data to evolutionary conclusions. Moreover, I will formulate a way to collect evidence in a more controlled way, in a laboratory environment, and present two studies that allow us to give a more detailed description of the workings of semantic protolanguage. At the end of this dissertation we will have a more well-founded and more detailed picture of semantic protolanguage, that shows that semantic organising principles are the evolutionary precursors of syntactic rules.

The hypothetical stage sketched above triggers some questions. For example: is it at all possible to talk about intermediate stages in the evolution of language (i.e., did language not appear abruptly)? What are the alternatives to a view in which semantic principles played a role in the emergence of language? The remainder of this chapter has as a goal, first of all, to provide answers to these questions. At the same time, I hope to familiarise the reader with the language evolution debate and specify my place in this debate.

I will, first of all, sketch the aspects of language that can be studied where the emergence of language is concerned: I will give a very brief overview of the subdisciplines of linguistics and their roles in the language evolution debate in section 1.2. Subsequently, in section 1.3 I will discuss three well known controversies in the language evolution debate and show how these controversies can be resolved, once we drop the assumption that there is one property of language that is a 'core' property, and adopt the view that language is a complex phenomenon, a composite system of many different capacities working together. I will show that this view on language leaves open the possibility of so-called *protolanguage* stages in the emergence of language. Section 1.4 presents different accounts of protolanguage and compares them to each other.

1.2 The sub-disciplines of linguistics

The human capacity to use language for communication is a complex skill, and it requires various abilities: encoding a mental representation in a linguistically meaningful string; pronouncing that string correctly; perceiving, decoding and

interpreting strings uttered by others. Many subdisciplines are involved in the study of this capacity: phonetics, phonology, morphology, syntax, semantics, pragmatics. Above, we have seen that I am especially interested in the role of semantics in the emergence of language, but all subdisciplines play a role in the language evolution debate. In what follows, I will briefly sketch how.

Semantics is the study of meaning in language. It is generally recognised that we know many words, and that all these words have meanings. Up until recently, it was not often discussed in much detail, however, how the meanings of words are combined, via complex syntactic rules, to form *structured meanings,* and how this property of language came about.

Fortunately, recent publications have started to explore the evolutionary history of human language with meaning in central focus (Hurford, 2007; Jackendoff, 2002), or with an emphasis on the origins of *compositionality* (this is a principle that tells us that the meaning of a complex whole is determined by the meaning of its parts and the way in which they are put together) (Smith *et al.*, 2003; Kirby, 2000).[2]

The aspects of **phonetics and phonology** are central when the origins of speech are studied. See e.g. de Boer (2001), where the origins of human vowel systems are studied, and section 3 in Fitch (2010), where the origins of the human vocal tract and phonological structure in human language are discussed.

Researchers focusing on the emergence of **morphology** ask themselves how human language got morphological structure, i.e., internal word structure. There are two prevailing views. One is that morphology originated from syntax: free elements in sentences became affixes. But this scenario cannot account for every form of morphology (it cannot explain alternations like *goose/geese* in English), and a second view is that morphological structure already emerged early, and came into existence through phonological alternations in speech (Carstairs-McCarthy, 2005). According to this view, stem alternations (e.g., *goose/geese* or *run/ran* in English) existed before affixes came into existence. The emergence of morphology is addressed extensively in Carstairs-McCarthy (2010).

The subdiscipline of **syntax** has received much attention in the language evolution debate. A reason for this might be that in a very influential publication, Hauser, Chomsky and Fitch (2002) hypothesise that the key step in the evolution of language is a syntactic step: the emergence of recursion, which makes more complex syntactic operations possible.

One of the topics that has been explored intensively is the structure of animal communication systems in terms of the Chomsky hierarchy, or formal language theory. In formal language theory, languages are seen as sets of sentences, and the rule systems generating these sentences can be arranged in a mathematical hierarchy. For example, languages that are recognised by a finite state grammar (*finite state languages*) are less complex than languages that

[2]A detailed overview of the discipline of semantics will be provided in chapter 3.

are recognised by a context free grammar (*context free languages*).[3]

The idea that 'evolution has climbed this hierarchy'[4] (in other words: the more complex the organism, the more complex the classes of languages it can recognise) has been appealing to many researchers. In Fitch and Hauser (2004), for example, it was shown that tamarins were able to systematically recognise strings from a finite state language (the strings used in the experiment were generated according to the rule (ab^n), leading to expressions of the form of, e.g., *abab* or *abababab*), whereas they failed to do this for strings from a context free language (the strings used in the experiment were generated according to the rule (a^nb^n), leading to expressions of the form *aabb* or *aaaabbbb*). Humans in contrast were able to recognise languages from both classes.[5]

This kind of experiment has received criticism; see e.g. (Zuidema, 2005, p. 123–124) and (Hurford, 2011). The latter reports research with birds, where it is found that some birds *do* recognise context free languages (see the study with starlings in Gentner *et al.* (2006)). Both Zuidema and Hurford note that the ability to recognise strings from a language with a certain complexity is hard to measure when these strings *have no meaning* (Zuidema, 2005, p. 124), (Hurford, 2011, p. 59).

In linguistics, when syntax is studied from within the Chomskian paradigm, natural language *meaning* plays only a marginal role. In the language evolution debate, however, concentrating on syntactic structure alone (divorced from meaning) seems no longer the dominant approach, as noted in Johansson (2005):

> Syntax, in the Chomskian generative paradigm, is an autonomous system totally decoupled from questions of meaning and function. But even though the sleep of colorless green ideas is as grammatical as it is furious, not a few linguists feel that the Chomskian 'syntactocentrism' (Jackendoff, 2002) may be a mistake. Other aspects of language cannot be neglected, and something central in language is missing when syntax is divorced from meaning. (p. 9)

A similar attitude towards structure and meaning, but in a slightly different context, can be found in Fitch (2010), in a discussion about the 'syntax' of birdsong. Fitch notes that the structure of birdsong should be seen as a phonological phenomenon, and not as a syntactic one, *because* birdsong does not have meaning:

[3] For a full description of the Chomsky hierarchy and formal language theory, and its role in the language evolution debate, I refer the reader to (Hurford, 2011, chapter 1).

[4] (Zuidema, 2005, p. 120)

[5] Everaert and Huybregts (2012) note that Chomsky distinguished between a 'weak' and a 'strong' generative capacity (respectively, the generation of correct strings, and the generation of correct strings with the right hierarchical strucutre). This distinction is often neglected in the literature about the emergence of language, also by Hauser and Fitch, but it adds complexity to the discussion about evolution 'climbing the Chomsky hierarchy'.

> Structure in [birdsong] is best treated as phonological (or musical), because there is no evidence that these vocalizations convey complex propositional information: they are not meaningful sentences. (Fitch, 2010, p. 183)

Although Johansson's quote above is—perhaps—a bit polemical, he does signal an important trend in the field, and I would like to follow the direction suggested: to no longer divorce syntax from meaning and to look at the origins of complex, compositional meaning. This might be the place where the roots of syntax are to be found.

To conclude, language is a complex phenomenon, and if we want to arrive at a complete account of its evolutionary roots, we need all subdisciplines of linguistics to contribute. The brief overview provided here has shown that until the moment we provide a full and complete account of the emergence of language, the different subdisciplines can have quite different goals. Sometimes, however, the subdisciplines can get in each other's way: when syntax gets too much emphasis, this may happen at the expense of semantics. When this happens, we might miss out on an elegant way to explain the origins of syntactic complexity, which would be a shame. In other words, it is time to look at the origins of language from the point of view of semantics. I will do this, by introducing and discussing a semantic account of protolanguage.

But before we can discuss *any* account of protolanguage, I need to make a reasonable case for why there could have been something like protolanguage. I will do this by making a small detour along some of the well known controversies in the language evolution debate. I will show that these controversies can be made less problematic once we concede that language is a 'bag of tricks' rather than a 'monolithic whole'. In other words, we should not try to treat language as one thing, or even define one aspect of language as a core or central property. Once we adopt a multi-component perspective on language, we automatically create room for protolanguage stages in its evolutionary history.

1.3 The many-components view of language

1.3.1 The many-components view resolves contradictions in the language evolution debate

Because language is such an important part of what makes us human, it is easy to have strong convictions about its nature and perhaps about its origins. It is not surprising that language evolution research knows many strongly debated issues. I will discuss three such issues here, and show that they need not be controversial, once one thinks twice about how language should be defined. The controversies I will discuss are the following:

- **adaptation vs spandrel:** does human language have a function, or is it rather a by-product of evolution?

- **gradual vs sudden:** was the emergence of language a gradual process or a sudden step?
- **innate and genetically determined vs learned and culturally determined:** is the human capacity to use language a largely natural or cultural phenomenon?[6]

For each of these issues, I will sketch the opposing positions. Throughout my exposition, I will show that one central idea offers us a way to solve the controversy in many issues: this is the idea that *language is a phenomenon that has many different subcomponents* (these components being general cognitive abilities like 'inference', 'Theory of Mind' and 'signal learning') and it should not be seen as an unanalysed whole.

The idea that language is a phenomenon that has many subcomponents not only offers a good way to analyse the controversies listed above, but it also makes it possible to take the idea of assuming an intermediate stage in the emergence of language (protolanguage) seriously. This step will be clarified below, in section 1.4.

Let me first explain the many components view of language and its background, before I continue to discuss the controversies mentioned above.

In linguistics, it has long been customary to stress the extent to which human language is special and unique to humans. The cognitive and communicative abilities of animals and their similarities with human abilities were not seen as relevant by many linguists, until in 2002 a paper appeared in Science by Hauser, Chomsky and Fitch (Hauser *et al.*, 2002). They argue that it is no longer useful to speak of the human language faculty as if it is an unanalysed whole (the internal component of the mind/brain that is sometimes called *I-language*). Instead, they hypothesise that the human faculty of language can be seen as divided into two parts: FLB (the faculty of language in a broad sense) and FLN (the faculty of language in a narrow sense). The former, they postulate, includes many mechanisms that humans share with animals, and only the latter is uniquely human. They further hypothesise that *recursion* is the key property of FLN, and they propose that this hypothesis should be tested empirically.

After the publication of (Hauser *et al.*, 2002), many researchers focused on FLN only, making FLN into 'core' language. The large body of publications focusing solely on the role of recursion in language evolution is symptomatic of this. But in a way, scholars doing this were again attempting to define language as something narrow and specific. To sum up, the introduction of the terms FLN and FLB into the language evolution debate has certainly served a purpose: it has invited many scholars, and linguists in particular, to start thinking about the emergence of language. However, the sole focus on FLN and recursion as the key step in language evolution has made the debate quite rigid.

[6] The issues described here and the terminology to describe them were taken from Johansson (2005).

An attitude towards language that is much more appealing, and that can help to solve many apparent issues in the language evolution debate, is the one adopted in a recent book by Tecumseh Fitch (Fitch, 2010), where he argues that FLB is the interesting phenomenon to study, and we should not look for a definition of language in which only one aspect of language is singled out:

> [L]anguage must be viewed as a composite system, made up of many partially separable components. Many of these components are widely shared with other animals (such as the capacity for hearing, memory, basic cognition, and vocalization), but a few differentiate humans from our nearest primate cousins (such as vocal learning or complex syntax). Crucially, each of these necessary components of language may conceivably have its own evolutionary history, and rely upon quite separate neural and genetic mechanisms. Although language is a system characterized by seamless interaction between these multiple components, "Language" is not a monolithic whole, and from a biological perspective may be better seen as a "bag of tricks" pieced together via a process of evolutionary tinkering. To the extent that this multi-component perspective is correct, any attempt to single out just one aspect of language as "core" or "central" is a mistake. (Fitch, 2010, p. 5)

Fitch calls this view of language the multi-components perspective, and from this perspective, many issues about which heated debate exists, e.g., about innateness or about the question whether language is an adaptive trait or a by-product of evolution, can be resolved or at least reformulated. Let us have a look at these issues now.

1.3.2 Controversy 1: adaptation vs. spandrel

In evolutionary biology, a distinction is made between structures that are adaptive and structures that appeared as unselected byproducts, 'spandrels'. In the language evolution debate, both properties are ascribed to human language. Bickerton (1998) and Chomsky (1988) consider language to be a spandrel: they see language as a by-product of evolution, on which natural selection had no influence. Pinker and Jackendoff (2005), on the other hand, defend the view that language is an adaptive phenomenon.

The question whether human language is a spandrel or an adaptation presupposes a view of language in which language is one thing: an unanalysed whole. As we have seen above, this is not a fruitful way to characterise human language. Once we take language to be a complex trait, consisting of many sub-capacities (examples of such sub-capacities are the capacity for referring to objects in the world, Theory of Mind, meaning composition, embedded structure, etc.), we can hypothesise that some of these components are adaptations (having been selected for some useful function), while others appeared as spandrels. At least, this is most likely for occurrences of complex behavior, also in

other animals. Thus, instead of quarreling about whether language is adaptive or not, we can concentrate on subcapacities of language and make the discussion more detailed and specific.

Apart from the terms 'adaptation' and 'spandrel,' there is a third term that is important, namely that of 'pre-adaptation'. This concept was already mentioned by Darwin, and a pre-adaptation is a trait or organ that was originally used for some function and was put to some new use in the course of evolutionary history. An example of a pre-adaptation is a gill (a respiratory organ found in aquatic animals). Gills were originally used by sea animals for breathing in water, but when, in evolutionary history, aquatic animals became air-breathing and semi-terresteral (and gills were no longer needed for breathing), they were put to a new use as, among other things, the larynx (Fitch, 2010, p. 63–64).

Function shift is not only common in morphological evolution (the evolution of bodily organs), but also played an important role in the evolution of cognition, as was claimed by Gould (1991). In other words, I suspect that there is at least some truth in the claim that *language is a new machine that Nature built out of old parts* (Bates 2003, translated and quoted in (Johansson, 2005, p. 167)).

To conclude, the question whether language as a whole is an adaptation is not the right kind of question. Instead, I take the human linguistic capacity to be a complex capacity, some subparts of which might be adaptations, and others might be spandrels. Moreover, I suspect that pre-adaptation played a role in the emergence of language. However, there is no conclusive evidence yet about the true nature of language concerning the evolutionary role of its subparts, and I will leave the discussion at this point. The most important observation is that it makes little sense to try to claim that language *as a whole* is either purely adaptive or a spandrel.

1.3.3 Controversy 2: gradual vs. sudden

Another issue that has been discussed intensively is the question whether humans acquired their capacity for language in one single step, or gradually. The former position was defended in Chomsky (1988) and Bickerton (1990), but nowadays, more and more researchers endorse the contrary view: that *de novo* evolution of many genes that work together is biologically highly unlikely (Johansson, 2005, p.170–171).

Also here, it is useful to recognise that the human capacity for language is not one unanalysed trait, but consists of many sub-capacities. Then it becomes possible to look at the genetic origins of all these subparts, and for these subparts it makes sense to speak of 'sudden' phenotypical changes, especially where old structures are put to new uses (pre-adaptations, or exaptations). In other words, sudden changes are not ruled out, but they act only on a relatively small scale. Moreover, as is pointed out in Fitch (2010), we should not overestimate the role of sudden genetic change in the evolutionary origins of humans:

> There is no *a priori* reason to reject a refined saltationist hypothesis of a relatively large phenotypic mutant playing a role in human evolution (the only variant of the non-gradualistic world view left standing). Importantly, however, there are no compelling examples of such changes at present. (Fitch, 2010, p. 55)

A similar position in favor of gradual emergence of language is formulated in Johansson (2005):

> [G]iven the near-impossible odds against the single-step appearance of something as complex as language, we can conclude that the evolution of language is overwhelmingly more likely to have been gradual, in the sense of entailing many small evolutionary steps, rather than a single leap.

In sum, there is no conclusive evidence whether language emerged gradually or relatively quickly, but to claim that language emerged in one giant leap is a very strong claim that has become very hard to maintain.

1.3.4 Controversy 3: innate and genetically determined vs. learned and culturally determined

The question whether language is innate or not is tightly connected to the language evolution debate. If one supposes language to be largely innate, then apparently, much of language has to have a biological explanation: it must be explained how the machinery necessary for language learning got into place and which genetic changes have been responsible for the emergence of such a complex trait. If, on the other hand, one takes language to be largely a cultural phenomenon, there is not such a biological burden of proof, but it should be explained how cultural processes could have accounted for a complex system like natural language.

The innateness question of language has been the subject of very fierce debate. I do not wish to go into very much detail about this debate, because its participants take the issue very much as a simple black-or-white dichotomy. My position is in the middle ground: I agree with a growing body of scholars that 'strong innateness' is not right, but I am not convinced that language acquisition is just *tabula rasa conditioning*. Also here, it makes sense to see the human capacity for language not as an unanalysed monolithic phenomenon, but as a complex whole. And this form of complex behaviour is guided by both genetic dispositions and environmental input (Fitch, 2010, p. 31).

Thus, language *learning* is very likely to be a result of biological as well as cultural processes. Similarly, the view that also the *emergence* of the language learning device (or faculty of language, if you wish) might be a result of both biological and cultural influences is becoming more and more widely accepted (see, e.g., Kirby *et al.* (2007), Fitch (2010), Scott-Phillips and Kirby (2010)).

The idea behind this is that not only are our learning skills well-adapted to learn language, but also that language is well-adapted to be learnt by us.

This view of the emergence of language will play a role below, and in later chapters, so let us have a brief look at it. The structure of language, according to this line of thinking, is not only the result of biological evolution. Thus, it is not only genetic change in humans that has made it possible for us to learn language, but a different process plays a role as well: *cultural evolution*. The thought behind this is that when a communication system with some complexity is in place, it has to be learned by individuals. The way these individuals learn it is partly with the help of their language learning device and partly by observing utterances made by others. When we assume that the language learning device is not the only factor that influences language learning, then these utterances of others are an important factor in the process: Individual learners who have acquired a language will pass this on to their offspring by producing utterances that are observed by their offspring. If an individual learner changes the language, or makes mistakes, he will produce different utterances and the changes or mistakes will pass on to future generations (who will again learn language by observing the utterances of their parents, or other members of previous generations). In other words, changes in the language are passed on through generations. It has been shown in computational and laboratory experiments that these changes tend to make languages more learnable (see section 2.3).

Language is thus seen as a system that can undergo changes over generations. In other words, language itself is something that has evolved. Aspects of language that are not easily learnt will not be picked up by future generations, and thus, language becomes adapted to learnability. This process is generally called *cultural evolution* or *glossogeny*; these concepts will become important below, in section 1.4.1 and 2.3.

1.3.5 Early stages in the evolution of language: protolanguage

Above, a picture of language was sketched that presents language as a broad phenomenon that consists of many sub-capacities, many of which are probably shared with animals. It was shown that once this *many components view* of language is endorsed, controversial issues can be resolved or re-analysed in a constructive way. In short, it is not right to simply ask: 'is language innate' or 'is language an adaptation', because that is simply not the right way of looking at language. Also for questions like 'did language emerge gradually or suddenly' and 'is language organised in modules', extreme answers are not likely to be the right ones. Given that language is complex and has many subcomponents, a sudden emergence of such a system is not likely to have taken place. It is much more likely that language emerged gradually. Of course, there might have been sudden lapses in this process, though no evidence for this has been found yet.

Now that we have established language as a composite system, we can start to think about earlier stages, in which language had not developed into the complex system it is nowadays. In other words, we have paved the way for an investigation of evolutionarily early language, or, as it is called in the literature, protolanguage:

> Since protolanguages constitute hypotheses about what a system could have been like, before it was linguistic, the very notion of a protolanguage requires that we abandon preconceptions about one "core" or central aspect of language. An open-minded attitude towards different hypotheses about protolanguage thus goes hand in hand with the multi-component approach. (Fitch, 2010, p. 10)

In the next section, I will reflect on different hypotheses about protolanguage, one of which is the semantic protolanguage that will be central to the remainder of this dissertation.

1.4 Protolanguage

1.4.1 The place of protolanguage in the emergence of language

What is protolanguage? The simple answer to this question is 'an intermediate stage between *the systems of thought and communication present in the Last Common Ancestor* and *modern language*' (Fitch, 2010). This simple answer is, by the way, already quite complex and potentially problematic: because why would we suppose that there was only one (relevant) intermediate stage between no language and full language? And if there were more, how do we distinguish between them (Smith, 2006)? Moreover, the quick definition takes for granted that we have clear definition of 'modern language' and of 'the Last Common Ancestor'.

But I am not going to be discouraged by these immediate questions and take comfort in the observation by Tecumseh Fitch that...

> the necessity for at least one intervening protolanguage stage in hominid evolution is nearly universally accepted today.(Fitch, 2010, p. 400)

There is not much consensus about what protolanguage should have looked like and this section will give an overview of different hypotheses about the protolanguage stage.

1.4.2 Accounts of protolanguage: lexical, gestural, musical, holistic

In Fitch (2010), four different accounts of protolanguage are described: lexical, gestural, musical and holistic protolanguage. I will briefly describe each, and sketch how these different accounts relate to each other.

Lexical protolanguage is what Fitch takes to be the view defended by (among others) Bickerton.[7] Bickerton (1991) poses the claim that there is a 'mode of linguistic expression that is quite separate from normal human language' (Bickerton, 1990, p. 122). Examples are the simple sentences observed in phenomena like pidgin languages ('Me no lie,' 'flour expensive') and the language of children under two ('me toothbrush,' 'mommy lunch'). Bickerton puts forward the hypothesis that this mode of communication, which he calls protolanguage, was an intermediate stage in the emergence of human language, and that the sentences observed in the phenomena mentioned above are living fossils of protolanguage.[8]

Bickerton pictures protolanguage as a stage in which there was a lexicon with meaningful words, but no complex syntax. Thus, this account of protolanguage paints a syntax-final picture of the emergence of language: it hypothesises that complex syntax emerged late in the evolution of language. An advantage of this view of protolanguage is that it offers an explanation for the existence of simple language systems (such as pidgin language and child language). A problem with Bickerton's conception of protolanguage is that he postulates a big leap from protolanguage to full language, in which syntax emerged catastrophically. Recently, more and more authors agree that this does not offer a satisfactory explanation for the emergence of complex syntax (Fitch, 2010, section 12.11).

A **gestural** account of protolanguage postulates a stage in the emergence of language where manual communication played a large role. Proponents of a gestural protolanguage, such as e.g., Corballis (2002) and Arbib (2005), point to the flexibility of gestures (they can be simple, but also quite complex), as well as the presence of gesture in our everyday language use (speech-accompanying gestures, but also emblematic gestures, such as the two extended fingers forming a peace sign) as evidence for a gestural protolanguage stage. Gestures are capable of fulfilling a linguistic function, and they are still present in our everyday behaviour, and this could indicate that gestures were important in human behaviour in our evolutionary history. Moreover, apes are quite good at acquiring relatively complex gestures, indicating that gesture was among the capabilities of the last common ancestor of humans and chimps. The existence of full sign languages, however, poses a challenge for proponents of gestural sign language. If it is possible to have a fully fledged language in the manual domain, why

[7] Fitch also presents Jackendoff as a proponent of lexical protolanguage, but below we will see that it is better to call Jackendoff's account of protolanguage *semantic* protolanguage.

[8] A more extensive description of these 'living fossils' will be provided below, in section 2.3.3.

then did we end up using speech instead of gesture? It is quite probable that gesture did indeed play some role in our evolutionary history, but it is hard to maintain that there has really been a gesture-only stage.

Holistic protolanguage, proposed by Wray (1998), postulates a stage where there was both a complex phonological system and a complex conceptual system, which were linked to each other without compositional mapping. These meaning/utterance pairs were transformed into a compositional language via a process of segmentation. As an example, Wray introduces two imaginary utterances, /mɛbita/ and /kamɛti/, which are associated with the meanings *give her the food* and *give me the food* respectively. After analysing the analogy between the shared syllable /mɛ/ and the shared meaning 'singular female recipient', speakers of holistic protolanguage will take this regularity into his inventory (Wray, 1998, p. 55). With segmentation happening bit by bit, phonological units without an ascribed meaning will be omitted and a compositional language is the result.

An advantage of the holistic account of protolanguage is that it offers a neat explanation for the existence of formulaic phrases (e.g., *How do you do?* or *Could you pass me the ...*). But holistic protolanguage has been subject to much criticism. One of the points of criticism is that it is cognitively demanding to re-analyse speech strings into regular patterns, and that this cannot be expected from early hominids (Tallerman, 2007). Another argument against the holistic protolanguage hypothesis is that nowadays humans put words together much more often than they take words apart (Heine and Kuteva, 2007).

A **musical** account of protolanguage is described and defended in (Fitch, 2010, chapter 14).[9] It postulates a stage in which meaningless song-like language was the main communication system among our evolutionary ancestors. This stage, Fitch proposed, is followed by a stage in which 'songs' were associated with meanings in a holistic way: structured strings and complex meanings were associated with each other in an arbitrary way (thus in the sense of holistic protolanguage, as described above). Subsequently, these complex strings are broken down into parts and these parts are associated with meanings, so that a compositional language system emerges, driven by cultural transmission. Fitch's musical protolanguage was proposed quite recently, and many of the hypotheses still have to be rooted in empirical evidence. The proposal generates predictions about the relation between language and music (e.g., the expectation that there is overlap between phonological and musical abilities within individuals), some of which are now being investigated.

An advantage of Fitch's musical protolanguage is that it offers an explanation for the emergence of complex vocal control in humans: it postulates that the pronunciation of complex music-like utterances occurred early in our evolutionary history. Moreover, it leaves space for other accounts of protolanguage

[9] Fitch discusses previous proposals of a musical protolanguage by Mithen (2005) and Brown (2000), and criticised by Botha (2009a). In this overview I will keep to the more recent proposal by Fitch himself.

to contribute to a full picture of the emergence of language. This multi-faceted picture has a disadvantage at the same time: an elaborate picture like this needs quite a body of empirical evidence to justify all the steps. Moreover, the crucial step in the proposed trajectory, the *musical* step, is slightly puzzling to me. It is proposed that in the musical protolanguage stage, meaningless song was the main *communication* system. Is this supposed to be a mechanism similar to that of birds, which also produce structured songs without meaning? Why, then, are human men and women both endowed with language, whereas birdsong is mainly a male business?

1.4.3 Semantic protolanguage

A semantic/pragmatic account of protolanguage is described in the work of Ray Jackendoff ((Jackendoff, 1999) and (Jackendoff, 2002, chapter 8)), and is partly based on Bickerton's claims about protolanguage.

As described above, in (Bickerton, 1990), the idea was introduced that we view linguistic phenomena like the language of under-twos and the language of trained apes as 'living fossils' of protolanguage. Bickterton's idea was worked out further, and illustrated with more linguistic data in (Jackendoff, 2002, chapter 8). Jackendoff follows Bickerton's main idea: the idea that protolanguage was an intermediate stage in the evolution of language, that this mode of communication is still present in our brains, and that it surfaces when modern language is disrupted. Jackendoff adds to this the observation that in a simple language system without consistent syntactic rules, one can still use the linear placement of words in an utterance in order to convey relations between these words.

The organising principles behind utterances are then semantic in nature and two examples of such principles are AgentFirst and FocusLast.

AgentFirst is a principle that says that the NP referent with the highest control comes first.

FocusLast is a principle that says that the information that is in focus, new information, should be at the end of the utterance.

Jackendoff points out that exactly these principles can be seen at work in phenomena that can be called living fossils. One of these is the Basic Variety. In the process of acquiring a second language outside the classroom, adult learners go through a stage that has been characterized as being (1) determined by a small number of organisational principles, (2) largely independent of the source or target language of the learner and (3) simple but successful for communication (Klein and Perdue, 1997).

The BV is a system that is put to use in a situation where modern language is disrupted. The organisational principles that play a role in it are exactly principles like AgentFirst and FocusLast.

Jackendoff's line of arguing can thus be summarised as follows:

- In simple language systems, the linear placement of elements in utterances can be used as a way to express relations between them.
- This way of structuring utterances is semantic/pragmatic in nature.
- These semantic/pragmatic principles are observed in situations where modern language is disrupted, and thus must have played a role in protolanguage.

Unlike Bickerton, Jackendoff does not endorse the view that the step from protolanguage to 'modern language' was one single miraculous leap, and instead speaks of a trajectory that takes us from protolanguage to full modern language.[10]

To sum up, Jackendoff observes that language systems without full-blown syntax do have organising principles, and that these principles are semantic in nature. He hypothesises that these principles played a role in protolanguage. In his incremental picture of the emergence of language, the ability to reflect semantic roles in word order came before full syntactical rules. In other words, structuring according to semantic or pragmatic properties stood at the basis of syntax. To me, this is a very appealing picture of language evolution. In what follows, I will refer to Jackendoff's hypothetical scenario as **semantic protolanguage** and I will further investigate this account of protolanguage. Note that among the principles that govern protolanguage, according to this account, are both semantic and pragmatic principles (e.g., the FocusLast principle could be called pragmatic), but the boundaries between semantics and pragmatics are not always clear and philosophy and linguistics know a long history of debate about how semantics and pragmatics should be defined (Bach, 1997). Within the language evolution debate, Hurford (2011, chapter 6.3) offers a relatively straightforward definition. He defines semantics as *sentence meaning*, and pragmatics as *speaker meaning*. However, it will become clear in chapter 3 that my use of the term semantics in this dissertation is broader than 'sentence meaning' alone. For this reason, I will continue to use the term 'semantic protolanguage,' but it should be kept in mind that this is a relatively broad definition of semantics.

One of the goals in this dissertation is to find out if there is further evidence supporting a semantic protolanguage hypothesis. On the basis of this evidence, I will specify in more detail which mechanisms are at play in such a protolanguage.

1.4.4 Holistic versus synthetic (or compositional) protolanguage

One opposition in the debate about the nature of protolanguage that needs special attention is that between holistic and synthetic protolanguage, because this

[10]Note that Bickerton himself, in more recent publications, has also shifted his position in favour of a more gradual emergence of language. See Calvin and Bickerton (2000).

has been discussed thoroughly in recent literature. The synthetic protolanguage hypothesis can be summarised as follows: language went from structurally simple to structurally complex. More specifically, single words arose first, and were combined later in evolutionary history, as syntax evolved (Tallerman, 2007). The account of protolanguage that will be central to this dissertation, semantic protolanguage, could be described as a specific instance of synthetic protolanguage.

Synthetic protolanguage is called compositional protolanguage in more recent literature (Arbib and Bickerton, 2010), and it is contrasted with holistic protolanguage (which was described above in section 1.4.2). In the debate between advocates of the two views Tallerman's critical analysis of holistic protolanguage (Tallerman, 2007) takes a rather central place. Part of Tallerman's criticism concerns the analysis that is required when multi-syllable words are decomposed into compositional structure. She argues that this task was too difficult, partly because of variability in language use: small distinctions between signals could have been considered potentially significant, and the analysis of holistic utterances into compositional ones could have been blocked by this. Another line of criticism she presents has to do with the kind of process that is postulated by proponents of holistic protolanguage: they claim that new words come into existence by a process of fractionation. But this process is not (or hardly) at play in fully fledged languages: "to propose a holistic strategy involving fractionation is to ignore the known processes by which words come into being in language" (Tallerman, 2007).

Proponents of protolanguage have taken up the challenge to refute the criticisms formulated by Tallerman, and have formulated challenges for a synthetic account of protolanguage that Tallerman proposes. Verhoef *et al.* (2012) present an iterated learning study with a set of whistled sounds. In this study, segmentation of holistic strings is possible because cultural transmission causes the inventory of sounds to become more predictable. They remark, however, that the results from their study show that combinatorial structure emerges so rapidly that a fully holistic protolanguage stage would not be very likely to have existed over a long time.

Smith (2006) points out that Tallerman has a simplistic account of the emergence of nouns and verbs, and of initial ordering constraints: she simply assumes that they were present in pre-human cognition. After pointing out this criticism, Smith (2006) proposes a way to unify both accounts of protolanguage: "a holistic protolanguage undergoes analysis to deliver up nouns, verbs, and some conventionalised ordering principles; the resulting synthetic protolanguage then feeds into known processes, such as grammaticalisation, to deliver fully modern language" (Smith, 2006, p. 321). He goes even one step further and suggests that a holistic and a synthetic stage were probably not strictly segregated; there could have been a stage where segmentation as well as composition took place (Smith, 2006, p. 322).

To summarise, semantic protolanguage and holistic protolanguage may ap-

pear to be strongly opposing views: one proposes that composition was important in the emergence of language, and the other claims that *de*composition had a central place. But these two views need not be mutually exclusive; there could even have been a stage in which processes of composition as well as decomposition played a role. How this would work exactly is an interesting question, but it is not a central concern in the context of this dissertation. Because I am convinced that semantics has not received as much attention as it deserves, I will focus on the role of semantics in the composition of elements in a protolanguage stage.

1.4.5 Protolanguage or protolanguages?

Before I continue to evaluate the properties of protolanguage, there is something I need to say about the term 'protolanguage'. The term was introduced into the recent debate by Derek Bickerton. When he did that, Bickerton had very clear ideas about the position of a supposed protolanguage stage in evolutionary history. It was an intermediate stage in which there was no syntax yet, and words were just put together. The step from protolanguage to full language, according to Bickerton, was one sudden leap. Thus, for Bickerton it was a very straightforward thing to claim that there was *one single protolanguage stage*: it was at the point in history where multiple words were communicated, but where there was no syntax at all, and there were no other stages in between this stage and 'full language'.

As we have seen, other researchers have picked up the term protolanguage, and hypothesised about its properties. But these other researchers are not necessarily proponents of the 'one huge leap' hypothesis of syntax (even Bickerton himself has dropped this view in more recent work), and tend to advocate a more gradual trajectory from no language to full language. But then, the place of protolanguage in evolutionary history is not straightforward anymore, and it is no longer sensible to speak of *the* protolanguage stage.

Should we abandon the concept 'protolanguage' altogether, then? Some authors seem to imply this, by calling protolanguage a speculative concept (Bidese *et al.*, 2012). I do not think that such hard measures are necessary, as long as we keep in mind that the term 'protolanguage' is more a tool for hypothesising about the emergence of language, than it is literally a point in time in the course of evolution. If this is kept in mind, the term protolanguage can help us to formulate more clearly hypotheses about how language came about:

> The protolanguage debate provides a fascinating test case for the development of evolutionary linguistics: it has the notable advantage that the opposing viewpoints are clearly stated, open to scrutiny, and pugnaciously defended. (Smith, 2010)

Thus, different accounts of protolanguage then simply emphasise different processes in the emergence of language. These accounts of protolanguage are

not necessarily mutually exclusive. If the possibility is left open that there was more than one protolanguage stage, then one could imagine that there could be truth in more than one of the accounts of protolanguage mentioned here. We have seen an example of this in Fitch's musical protolanguage scenario.

To summarise my attitude towards the different accounts of protolanguage, I suspect that there is truth in the idea that at some stage in the emergence of language, semantic properties played a role in the organisation of utterances, and thus formed a precursor to full syntax. This idea is embodied in the semantic protolanguage account, and in this dissertation I will investigate that account of protolanguage. I will investigate whether there is further evidence in support of a semantic account of protolanguage, and hypothesise—on the basis of this evidence—which mechanisms could have been at play in a semantic protolanguage. This does not mean that I exclude every other account of protolanguage; to me, the hypothesis that there have been several protolanguage-like stages (each with different characteristics) is still quite attractive. However, in the context of this dissertation I will not sketch a full trajectory of the history of language, and rather focus on semantics alone. I hope to make clear that the role of semantic properties in early language was greater than many people (including Fitch) suppose it was.

1.5 Conclusion and overview

This chapter had two goals: first of all, to specify and motivate my domain of interest in the broad domain of the language evolution debate, and secondly, to give the reader a brief introduction into what is going on in this debate.

The aspect of language I am especially interested in is that of semantics. I am not primarily interested in the meanings of individual words, but rather in meaning as a compositional phenomenon: the way word meanings are composed into complex, structured wholes.

It has been observed in the literature (Jackendoff, 2002) that in simple language systems, semantic properties of words determine their place in utterances. In other words, semantics drives syntax. From this observation, one can extrapolate that something similar once happened in the emergence of language: when language was still relatively simple, syntactic rules emerged out of semantic properties. I called this hypothesis about evolutionarily early language the *semantic account of protolanguage*. In the remainder of this dissertation, I will investigate this account of protolanguage. More specifically, the following questions have come up in the course of this chapter:

- Is there further evidence for a semantic protolanguage?
- What are the mechanisms in semantic protolanguage?

These questions will be central to this dissertation, and my aim is to answer them properly. I should warn the reader, though. Because language evolution is a topic that unifies many disciplines, the range of topics and methods that I will take up is quite broad. To give some examples, I will focus on linguistic data from adult second language learners, philosophical reflections about meaning and human cognition, animal behaviour studies, and semantic frameworks. Moreover, I will present a series of laboratory experiments. All this will contribute to making semantic protolanguage an account of protolanguage to be taken seriously.

Chapter 2 will take us from the questions, hypotheses and controversies that were identified in this chapter to ways to provide answers about the emergence of language. I will, first of all, sketch a series of possible ways to obtain (indirect) empirical evidence about the emergence of language. Subsequently, I will zoom in on the approach that investigates protolanguage by studying *restricted linguistic systems*. These systems, already described briefly in this chapter, are claimed to provide evidence for semantic protolanguage. I will provide an overview of this evidence, evaluate the approach, and list the issues that should be addressed in order to make a serious case for semantic protolanguage. These issues are connected to data collection (how can we collect data more easily?), as well as to the theoretical underpinnings (how can we make the step from linguistic data to evolutionary conclusions?).

We cannot make a case for semantic protolanguage, without having a clear picture of the phenomenon that is studied in semantics: meaning. **Chapter 3** will take a tour along philosophical, linguistic and evolutionary accounts of meaning. It will become clear that there are three possible intuitions about meaning. Firstly, our words and sentences refer to things in the world, or to abstract objects: propositions. Secondly, when a speaker makes an utterance, he typically intends his audience to believe something; in other words, meaning is dependent on the speaker's cognitive state. Thirdly, our utterances have the meaning they have because of linguistic conventions. These last two intuitions play a role in evolutionary accounts of meaning. I will describe two possible evolutionary roads to full language: one based on cognition; the other based on communication.

Chapter 4 takes up restricted linguistic systems where they were left off. It describes a way to extend the basis of empirical evidence for semantic protolanguage, by gathering data in a controlled setting, in a laboratory experiment called *the improvised communication task*. The second part of chapter 4 provides a theoretical underpinning for this approach, by formulating a bridge theory that employs and combines the cognition-based and the communication-based views on meaning defined in chapter 3.

Having defined a way to collect empirical data in the laboratory, and having formulated a well-founded way to draw evolutionary conclusions from the data, we will turn to actual laboratory studies in chapter 5 and 6.

Chapter 5 investigates how simple propositions are expressed in impro-

vised communication and extends the observation made by Goldin-Meadow *et al.* (2008) that there is one order in which simple events are represented, by showing that events with different semantic properties lead to different orderings. A second experiment shows that different orderings in improvised communication are interpreted differently, thus suggesting that using different orders really has a communicative function.

Chapter 6 investigates what happens in improvised communication when information about temporal location is added to simple propositional content. In other words, how do people communicate that an event took place at some time other than now, in a communication system that lacks fully developed grammar? It is shown that the event information and the temporal information are conveyed separately: temporal information is added to the periphery of improvised utterances, along the same lines as is observed in restricted linguistic systems.

Finally, **Chapter 7**, sums up the results of this dissertation: semantic protolanguage should be taken seriously as an account of protolanguage. Evidence for this account can be found in natural restricted linguistic systems, but this evidence is backed up by results from laboratory experiments. The experimental approach developed in this dissertation, improvised communication, offers the possibility to investigate precise hypotheses about the emergence of language in a controlled setting.

CHAPTER 2

Evolutionary evidence: restricted linguistic systems

2.1 Introduction

In the previous chapter I have introduced the questions that will be central to this dissertation. This chapter will focus on answers, and it will evaluate evidence supporting a semantic account of protolanguage. But before focusing on this particular kind of evidence, I will reflect on the collection of empirical evidence about the emergence of language in general.

Given the fact that the emergence of language in humans has left no direct evidence (how nice it would be to have sound recordings of our evolutionary ancestors!), scholars have started focusing on *indirect* ways to gather evidence. We have seen some examples of how evidence is used in the language evolution debate in the previous chapter, but in this chapter, ways to obtain indirect evidence will be discussed more systematically.

Indirect data like this can be acquired in the following ways:[1]

1. By looking at the behaviour of nonhuman animals, to draw conclusions about the systems of thought and communication of the last common ancestor of apes and humans (**the comparative approach**).

2. By modeling communicative situations in a computational setting to

[1]This set of approaches is an updated and adapted version of the overview of approaches described in Kirby (2007). The caution the author gives there also applies here: the areas listed do not give an exhaustive description of the field (for instance, it does not cover the disciplines of genetics and anthropology), but merely gives a flavour of the predominant issues at hand.

draw general conclusions about communicative interaction (**the computational modelling approach**).

3. by studying the communicative behaviour of participants in a lab situation, when they are asked to perform communicative tasks or tasks that involve language learning (**language evolution in the lab**).

4. By studying situations where people are not able to use their first language in a normal way and are forced to improvise, e.g. when a deaf child grows up in a hearing family that does not know any conventional sign language (**restricted linguistic systems**).

In this chapter, I will focus in detail on restricted lingusitic systems. The approach that studies these systems in order to draw evolutionary conclusions is also called the *windows approach* (Botha, 2009b). This approach can provide us with the kind of evidence to answer the questions formulated in the previous chapter: is there further evidence for a semantic protolanguage, and what are the mechanisms at play in semantic protolanguage? I will discuss data from various restricted linguistic systems and evaluate the status of evolutionary conclusions drawn from the data. The picture of the windows approach that emerges from this evaluation is one of an approach that is promising, but that can benefit from more data and further work on the theoretical underpinnings. The open issues encountered in this chapter will be addressed in later chapters of this dissertation.

Before I focus on the windows approach, I will give an overview of all approaches listed above. This will give the reader an idea of the kind of data that is studied in the language evolution debate, and it clarifies the position of the windows approach in the debate.

The ways of collecting data described above have in common that they look at something other than the emergence of language in order to draw conclusions about the emergence of language, and one huge advantage of this is that the focus in the debate is on empirical data instead of hypothesised scenarios.[2] Thus, now it is at least possible to falsify evolutionary hypotheses by reference to data.

A way in which the approaches differ from each other is related to the distinction between biological and cultural evolution that was introduced in chapter 1. The comparative approach investigates the differences between species, and thereby focuses mainly on *biological* issues. The other approaches, on the other hand, focus on language forms and interaction, thereby focusing mainly on *cultural* phenomena. It should be said, though, that these approaches have something to say about biological issues as well, but in a more indirect way.

In the next section (2.2), I will describe a way to collect evidence about the biological aspects of the evolution of language. Subsequently, in section

[2]This has not always been the case; see remarks in (Fitch, 2010, p. 16) on the bad reputation of the language evolution debate.

2.3, I will describe three ways to investigate cultural aspects of the evolution of language. Lastly, I will dig deeper into the windows approach: I will look at data from restricted linguistic systems (in section 2.4) and evaluate the approach (in section 2.5).

2.2 Studying biological evolution: the comparative approach

2.2.1 Investigating homologous and analogous traits

In the comparative approach, empirical data about the cognitive and communicative abilities of animals is studied in order to formulate hypotheses about which features of language are uniquely human. The comparative method was put to use even before evolutionary biology existed, but the statistical foundations of it were laid by Harvey and Pagel (1991). The comparative method compares traits among species, but this can be done in two ways. The first is to look at **homologous** traits: traits that were present in a common ancestor of the species under consideration. An example of such a trait is fur in mammals. The study of homologous traits can be used to reconstruct the last common ancestor of humans and chimpanzees. Thus, in the language evolution debate, the study of homologous traits is usually restricted to the study of apes.

A second way of applying the comparative method to study biological evolution is by looking at **analogous** traits: traits that are not related by descent and were thus not present in the common ancestor of two species, but evolved independently in two separate lineages. An example of a structure that evolved many times, in different species, is the eye. Studying the eyes of species in different lineages can teach us about the function of vision.[3] In the language evolution debate, scholars have recently taken up the study of birdsong, which can be seen as analogous to human language: in both, social learning plays an important role, and there also appear to be many similarities in the way birdsong and human language are represented in the brain (Bolhuis *et al.*, 2010).

The comparative approach has acquired a central position in the language evolution debate, partly because many scientists have abandoned the early Chomskian view that language evolved entirely *de novo* in humans. Important in this development was a publication by Chomsky himself (Hauser *et al.*, 2002). As we have seen in chapter 1, Hauser *et al.* (2002) propose to distinguish between the *faculty of language—broad sense* (FLB) and *faculty of language—narrow sense* (FLN). The former, they postulate, includes many mechanisms that humans share with animals, and only the latter is uniquely human.

Subsequently, they claim that the contents of FLB and FLN must be determined empirically rather than on an *a priori* basis and they put forward the hypothesis that, first of all, FLN essentially contains recursion, and secondly,

[3]See (Fitch, 2010, p. 46), and the references provided there.

only FLN is uniquely human. Thus, FLB shares much of its functionality and structure with animals. They present a body of empirical data that suggests that their hypothesis is right, but they invite people to continue the empirical investigation into the correctness of their claim:

> Linguists and biologists, along with researchers in the relevant branches of psychology and anthropology, can move beyond unproductive theoretical debate to a more collaborative, empirically focused and comparative research program aimed at uncovering both shared [...] and unique components of the faculty of language.
>
> Although we have argued that most if not all of FLB is shared with other species, whereas FLN may be unique to humans, this represents a tentative, testable hypothesis in need of further empirical investigation. (Hauser *et al.*, 2002, p. 1578)

Since then, many empirical studies on the communicative and cognitive behaviours of various animals have been incorporated into the language evolution debate. Fitch (2005), for example, gives an overview of comparative studies on, e.g., perception and interpretation in nonhuman primates: he reports that many monkeys react appropriately to alarm calls by conspecifics but also to alarm calls from other species (p. 205). Moreover, they are able to interpret sequences of calls (p. 205), and this leads him to the conclusion that 'nonhuman primates having quite complex minds' (p. 205).

In a review of the 8th International Conference on the Evolution of Language, Balter (2010) notes that the focus on empirical data has done the language evolution debate much good. Recent work that is mentioned in this review is a study on chimpanzee food calls (Schel *et al.*, 2010), which shows that free ranging chimpanzees were more likely to produce food-associated calls when individuals in the group that they groom ('chimp allies') were present, rather than absent. This suggests a relation between grooming and communicating.

The increasing attention to birdsong in the language evolution debate is reported as well in Balter (2010). An example of this is Ohms *et al.* (2010), who show that zebra finches were able to distinguish between two similar sounding Dutch words (*wit* and *wet*), even when the words were pronounced by different (male and female) speakers. They thus show evidence against the claim that only humans can distinguish between similarly sounding words.

2.2.2 Underlying questions

In Kirby (2007), an inventory of possible evolutionary questions is discussed, and four categories of questions are distinguished that underly language evolution research: questions about the uniqueness, structure, function and history of language:

Structure: Why is language the way it is and not some other way? How can an evolutionary approach explain the particular language universals we observe?

Uniqueness: Why are we unique in possessing language? Which features of the human capacity of using language are truly uniquely human?

Function: How could language evolve? What were the selective pressures involved?

History: What is the evolutionary story for language? When did it evolve? Were there intermediate stages? What did these intermediate stages look like? (Kirby, 2007, p. 5, adapted)

The four categories are quite different from each other and they reflect the different goals of scholars working on questions about language evolution. As observed in Kirby (2007), in many cases these questions are only asked implicitly and it might even be that confusion in the debate stems from the fact that researchers are asking different underlying questions. However, answers to the individual questions are not entirely independent; they influence each other. For instance, when one formulates a hypothesis about the uniqueness of a certain feature of human language, at the same time one says something, indirectly, about the structure of language. Let us now look at which questions underlie the comparative approach.

Work that is carried out within the comparative approach has, first of all, much to do with the question about the *uniqueness* of features of language, especially when one focuses on the features that are presumably part of FLN. But questions about *structure* play a role as well: finding that a certain property is present in animal cognition or communication tells us that we have something in common with that animal, but at the same time it gives us information about where that specific feature came from, and how it might have been used by our evolutionary ancestors. But questions about uniqueness and structure only play a role when *homologous* comparisons are made; in the case of analogous traits, the underlying questions are different.

When studying *analogous* traits, for example in the case of birdsong and human language, one is not interested in the properties of a common ancestor. Because the traits that are being compared evolved independently (this is called convergent evolution), and there is no ancestral relation between them, it becomes possible to answer questions about the *function* of such a trait. In some birdsong studies, like the one in Ohms *et al.* (2010), uniqueness questions play a role after all.

Answers to evolutionary questions about function and uniqueness will not contribute directly to questions about the role of semantics and pragmatics in protolanguage. However, when a semantic account of protolanguage is embedded in a full evolutionary scenario, these questions become relevant after all. We will see this in chapter 3 of this thesis.

2.3 Studying cultural evolution: computational modeling, language evolution in the lab, restricted linguistic systems

Looking at how animals behave is not the only way to collect indirect evidence about the emergence of language. In this section I will discuss three ways to collect indirect evidence that focus not so much on biological properties of species and their role in the emergence of language, but on cultural processes. These are the *computational approach*, the approach that studies *language evolution in the lab*, and the *windows approach*, which studies *restricted linguistic systems*.

2.3.1 The computational approach

One way to study the cultural dimension of language evolution is to use computational models in order to place constraints on hypotheses about language evolution. Populations of individual agents are modelled in order to study the interaction between individual learning mechanisms and the change of linguistic behaviour over time. It is assumed that each agent learns the rules of a language first of all on the basis of the setup of its language learning device, but also on the basis of the linguistic output of its 'parents' (an earlier generation of agents). On the basis of the rules an agent learns, it produces output himself, which is in turn the input for a next generation. After many generations, a chain exists, and the linguistic output in each generation in the chain is the input for the next generation. Differences in language use between the parent and the child lead to change in the language that is used. Things become interesting whenever there is imperfect information about the target language: an agent does not see everything of a language it is learning. In real life, the same situation exist for first language acquisition, and it is called *poverty of the stimulus*. If this is the case, language users change the language slightly, and the input that the next generation of language learners will get is different. After a number of generations, language will change and optimise for learnability. Thus, the structure of a language is (partly) the result of repeated learning by individuals who learn a language, not by having a fact sheet with grammar rules, but by encountering a limited set of example sentences in the language.[4]

One of the many examples of a model that shows how linguistic phenomena can emerge from repeated learning by agents is provided in Smith *et al.* (2003). Here, it is shown that individual agents learning a language who are not shown the full set of examples in order to learn the language entirely do end up communicating successfully. And, more importantly: gradually, over the gen-

[4]Zuidema (2003) describes this as 'the poverty of the stimulus that solves the poverty of the stimulus,' because this view on language explains how children that supposedly have insufficient evidence in order to induce grammar rules can still learn a language.

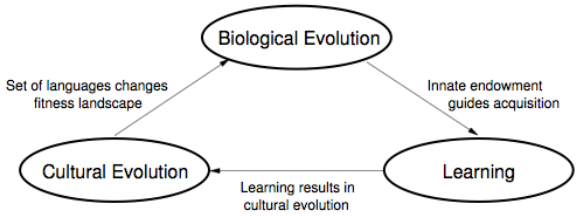


Figure 2.1: Three systems involved in the evolution of language, from Smith *et al.* (2003)

erations, these agents adapt the language by applying *compositional structure*. I.e., if a model starts with random names for objects, then the names will show systematicity after a number of generations, such that a word for a *green ball* will share properties (a syllable, for example) with the word for *green block*; thus, a syllable meaning, roughly, *green* comes into existence.[5]

This kind of computational modelling is called *iterated learning*: language is seen as a system that is 'repeatedly transformed from external linguistic behaviour to internal linguistic representation to external linguistic behaviour and so on' (Kirby, 2007, p. 10). A consequence of viewing language in this way is that it becomes possible to conceive of language itself as an evolutionary system. It is a system that can change and, under certain circumstances, will adapt towards becoming more easily learnable. This changing and adapting happens on a cultural level rather than on a biological level: it is not (only) the species itself that changes, but the cultural phenomenon, i.e., language itself.

When the possibility of cultural evolution is taken into account, the complete picture of the emergence of language becomes an interplay between biological evolution, cultural evolution and individual learning; see figure 2.1. In other words, human genes shape their bearer's individual learning mechanisms; these learning mechanisms play a role in shaping language; and finally, language influences the fitness landscape and thus affects biological evolution.

To sum up, the computational approach aims to show that some features of language can emerge from individual interactions among agents and individual learning by the agents. By doing that, it makes a claim about the *structure* of language (see the inventory of questions underlying language evolution research above in section 2.2.2).

The computational models that have been developed are important for this dissertation, because they show how the human capacity to learn complex structure, cultural interaction and biological evolution could have interacted with each other to produce a system as complex as human language. Moreover, the notion of iterated learning has recently been taken from the domain of compu-

[5]This example is to illustrate the principle; the model uses more abstract representations of meanings and utterances.

tational modeling and put to the test in a laboratory, using human instead of computational agents. I will describe this development in the next section.

2.3.2 Language evolution in the lab

Recently, iterated learning models have been conducted in the lab using human participants instead of computational models. In Kirby *et al.* (2008), a set of structured meanings was constructed such that each item had one of three different shapes, one of three different colours and one of three different ways of moving (e.g., a bouncing blue square). These meanings were paired up with strings of syllables (e.g., 'kihemiwi') and participants were asked to learn a language consisting of some of these meaning-string pairs; they were told that they were learning an alien language. After the learning phase (which had only a subset of the 27 possible items), they were tested on all 27 items (which means that they were tested on items they hadn't seen before), and the output of the first participant was used as the input for the following participant (whose output was used for the next participant, etc.).

After ten generations, the language had become more learnable, because only a few different words were used for the different items. In other words, there were many cases of homonymy. This loss of expressivity of the language was not random, however: the words underspecified the meanings in systematic ways. For example, one and the same word would be used for any bouncing object.

In a second experiment, where all cases of homonymy were removed from the learning set by the experimenters, the languages produced by participants became—over the generations—increasingly structured and began to exhibit compositionality: a string associated with a picture would consist of substrings representing the color, shape and movement of the item. Thus, the languages became both more learnable and more expressive. In other words, results obtained in computational models were successfully transferred to human participants.

This branch of the iterated learning paradigm has as an advantage that it can investigate the emergent properties of individual interactions using *actual humans* instead of computational agents.

The study described above shows how repeated learning can lead to regularities in a language. In this study, the language was already in place, and the participants were asked to learn it. In other experiments, it was investigated how people behave when they are asked to communicate in the absence of a communication system. In these experiments, not learning but *communicative interaction* is the central concept: how does repeated communicative interaction influence the structure of communication?

Garrod *et al.* (2007) investigated the emergence of a system of graphical communication by conducting experiments with a setup similar to the Pictionary game: pairs of participants were asked to communicate about concepts (such as *Clint Eastwood*, *cartoon* or *computer monitor*) by drawing pictures.

The task had several rounds, so the same concepts would occur several times during the experiment. Garrod *et al.* (2007) show that in several rounds of communicative interaction (drawing and guessing with feedback), iconic signs (signs containing many details) evolve into symbolic signs (signs containing less details).

Fay *et al.* (2010) compared two graphical communication tasks: one in which isolated pairs of participants interacted (like in the experiment described above) and one in which participants engaged in a task as a member of a community, in this case, where they interacted with seven different partners drawn from the same pool. Again, several rounds of drawing and guessing were carried out, resulting, eventually, in signs with less and less detail. It was shown that the symbolic signs that were the result of the isolated pair-wise interactions were as effective (in terms of communicative success) as the symbolic signs that were the result of the community-based interactions. Further, it was shown that interaction was crucial for the development of a *shared* sign system: separate pairs of participants ended up with different local sign systems, whereas communities of participants ended up with a single sign system shared throughout the whole community. These results are important to the iterated learning paradigm, because in this paradigm, individual learning is central, and one-way intergenerational interactions (children learning language from their parents) are seen as crucial to the development of language. By showing that interaction within a community (this is referred to as *horizontal interaction*) is important in order to arrive at a shared sign system, Fay *et al.* (2010) shed a somewhat different light on cultural evolution.

The method of studying language evolution in the laboratory is relatively new and it might still be too early to make generalisations about the experiments. Still, the experiments mentioned here show that repeated learning over generations, as well as repeated interactions *across* a generation, have an effect on the structure of newly emerging communicative systems. It has been suggested (Scott-Phillips and Kirby, 2010) that vertical cultural transmission (iterated learning) requires that a language is learnable, while horizontal interaction requires that a language is expressive.

The horizontal interaction approach will prove relevant for the experiments taken up in this dissertation (see chapter 4).

The approach discussed in the next section also studies data from modern humans in order to draw conclusions about the emergence of language, but within this approach data is not collected in the lab, but 'in the wild': it studies so called 'restricted linguistic systems'.

2.3.3 Restricted linguistic systems as living fossils

Suppose you are in a situation where you cannot use your native language or any of the languages you speak. In such a situation you have to improvise, using anything that comes up to get your message across: using gesture, pointing, or maybe sounds or an occasional word of a language you don't know well. In these

situations there is no opportunity to use sophisticated grammatical structures, and the kind of language that is used is the very core part, the very essential elements.

In section 1.1.3 I described the claim that the simple language systems that emerge in such 'improvisation' situations can tell us something about the emergence of language or, to be more specific, about protolanguage. This idea, to make the step from special linguistic situations to conclusions about the evolutionary history of language, was introduced by Bickerton (1990), and worked out further by Jackendoff (2002). In the work of Rudolph Botha (e.g., Botha (2009b), Botha (2006b)), linguistic phenomena that are studied in order to draw conclusions about aspects of language evolution are called *restricted linguistic systems*, and the approach that applies this method is called *the windows approach to language evolution*. Let me give a quick recap of Bickerton's and Jackendoff's ideas.

Bickerton (1990) introduced the idea that we can view linguistic phenomena like the language of under-twos and the language of trained apes as 'living fossils' of protolanguage. Bickterton's idea was worked out further and illustrated with more linguistic data in (Jackendoff, 2002, chapter 8). Jackendoff follows Bickerton's main idea: that protolanguage was an intermediate stage in the evolution of language, that this mode of communication is still present in our brains, and that it surfaces when modern language is disrupted. Jackendoff starts off with the general idea that in a simple language system without consistent syntactic rules, one can still use the linear placement of words in an utterance in order to convey relations between these words.

The organising principles behind utterances, Jackendoff claims, are semantic in nature, and two examples of such principles are AgentFirst and FocusLast.

AgentFirst is a principle that says that the NP referent with the highest control comes first.

FocusLast is a principle that says that the information that is in focus, new information, should be at the end of the utterance.

Jackendoff points out that exactly these principles can be seen at work in phenomena that are called living fossils. One of these is the Basic Variety. In the process of acquiring a second language outside the classroom, adult learners go through a stage that has been characterized as being (1) determined by a small number of organisational principles, (2) largely independent of the source or target language of the learner and (3) simple but successful for communication (Klein and Perdue, 1997). The Basic Variety is a system that is put to use in a situation where modern language is disrupted. The organisational principles that play a role in it are exactly principles like AgentFirst and FocusLast.

Jackendoff's line of arguing can thus be summarised as follows. First of all, in simple language systems, the linear placement of elements in utterances can be used as a way to express relations between them. Secondly, this way of

structuring utterances is semantic/pragmatic in nature. Thirdly, these semantic/pragmatic principles are observed in situations where modern language is disrupted, and thus must have played a role in protolanguage.

Jackendoff's ideas are interesting and can provide an interesting way to collect more evidence about the structure of protolanguage. However, there is still room to strengthen both the empirical and theoretical underpinnings of the 'living fossil' approach. Important work in this enterprise has been done by Rudolph Botha. In a series of publications he formulates conditions under which living fossils should be studied and interpreted, and introduces the term 'windows approach': the sources of evidence mentioned here can be used to provide a 'window' on language evolution.[6]

Central to Botha's foundational work on the windows approach is the demand to make claims about protolanguage (that are made on the basis of linguistic evidence) *systematic and specific*. The data that is used as a source has to be described insightfully, just like the evolutionary conclusions that are formulated on the basis of it. And most importantly, the step to go from the former to the latter has to be underpinned by a non-ad hoc *bridge theory*.

Botha not only stressed the importance of evaluating individual windows on language evolution, but also initiated a collection of papers in which data from various so called restricted linguistic systems are presented (Botha and de Swart, 2009): among others homesign systems, pidgin, unsupervised second language acquistion, but also compound nouns. Further, initial steps have been taken towards the formulation of bridge theories for some of the windows (de Swart, 2009; Roberge, 2009), and the combination of different windows to study the same phenomenon (Benazzo, 2009). In sections 2.5.2 and 4.3 I will focus in more detail on bridge theories. The next section will review data from various restricted linguistic systems.

In terms of the evolutionary questions formulated in section 2.2.2, what do restricted linguistic systems tell us about language evolution? The main underlying questions of this approach are, naturally, *historical* questions, as finding out what intermediate stages in the evolution of language looked like is central to this approach. Some other questions, however, are addressed indirectly.

For example, if one can show that there has been, at some point in our evolutionary history, a stable protolanguage stage that was successful for communication, then that answers questions about the *function* of language: it shows that at this particular stage, the function of language was likely to be communication.

To conclude, the windows approach is a growing and promising approach that offers linguists a valuable way to contribute to the debate. Moreover, it might help us to find answers to the questions I formulated in the previous chapter, about the emergence of structured meanings in language, and the relation between form and meaning in protolanguage.

[6]Botha discusses not only linguistic data as windows, but also archaeological data (Botha, 2003). The latter falls outside of the scope of this overview.

In the remainder of this chapter I will focus in more detail on restricted linguistic systems, and provide an overview of data from various systems and evaluate this approach to language evolution.

2.4 Restricted linguistic systems: data

In this section I will describe six restricted linguistic systems: the Basic Variety, pidgin, and three kinds of emerging sign systems (homesign, Al Sayyid Bedouin Sign Language and Nicaraguan Sign Language) . I will give a general description of each system.[7] This overview will give us a clearer idea of the claims that have been made about these systems, and it will allow us to evaluate the windows approach in the next section.

Within the six restricted linguistic systems, I will look at whether Jackendoff's principles AgentFirst and FocusLast are observed, but before we do that, I will focus on the Topic/Focus distinction as it has been defined in the literature, because this will help our understanding of the FocusLast principle.[8]

2.4.1 Topic and Focus

The terms *topic* and *focus* have to do with the way in which we structure information in language. We have seen the terms applied in chapter 2, but at this point I will give a brief overview of what has been said about topic and focus in the literature. Topic and focus can be seen as pragmatic phenomena (because they help a listener to interpret information), but also as semantic phenomena (because they can change the truth conditions of sentences in some cases). Below, I will discuss both effects.

Much of our communication takes place in a context, and interpretation of sentences is said to be *incremental*: a new sentence is interpreted in the context of previously uttered sentences. Therefore, many sentences can be analysed as consisting of two parts: one part connects to the information that is present in the context already (the information that is *given*, or *old* information), and another part specifies new information. In the literature, various different terms are used to describe this intuition (e.g., the distinctions 'topic-comment' and 'theme-rheme'), all slightly different from each other. In the overview below, I will follow the terminology used by de Swart and de Hoop (2000).

[7] It should be noted that Hurford (2011) gives a similar overview of restricted linguistic systems. In the book, however, these systems take up a different place in his general line of argument. Hurford claims that over the years, languages have become increasingly complex, and that the systems that we have called restricted linguistic systems represent 'growth rings' of language. In other words, Hurford claims that we can read off the structure of evolutionarily early forms of language directly from the simple language forms he discusses. My view on the way in which we can derive evolutionary conclusions from restricted linguistic systems is different; I will specify this view in section 4.3.3.

[8] I am assuming that the other principle, AgentFirst, is straightforward enough to do without a separate definition. However, section 3.8.3 will focus on the properties of the Agent relation.

In modern languages, there are different ways to encode old and new information: intonation, word order and morphological markers. The first two are employed in a language like English. To see how intonation marks new information, let us have a look at two examples, in which stress is indicated with capital letters (from de Swart and de Hoop (2000)):

(1) a. What does Susan want to drink?
b. Susan wants BEER

(2) a. Who wants beer?
b. SUSAN wants beer

In (1-b), 'Susan' is the topic and 'beer' is new information (focus); this element of the sentence is stressed. In (2-b), the words of the sentence are exactly the same, but this time, 'Susan' is new information and is therefore stressed.

A second way to give information about the informational status of elements in a sentence is varying word order. Consider the following two examples (again, from de Swart and de Hoop (2000)):

(3) a. When did Jane leave?
b. Jane left at six o'clock.

(4) a. What happened at six o'clock?
b. At six o'clock, Jane left.

In sentence (3-b), Jane is the topic, and the time at which she left is new information. Sentence (4-b) works very well as an answer to question (4-a). In (4-b), the point in time is the topic, and what happened at that point is the focus information. The following is an example of a cleft construction, a construction in which new or particularly interesting information (in this case 'beer') is put in the first position:

(5) It is beer Susan wants.

Word order variation to indicate differences in information structure in English only works in special cases, like the temporal example and the cleft sentence above. There are languages however (e.g., Hungarian and Czech) in which word order determines the informational status of clauses more clearly.[9]

To summarise, sentences can be divided into a part that connects to the information that is already known to the speaker and hearer, and a part that provides new information. These parts are called the topic and focus parts of the sentence, respectively. Full, modern languages have different means to express the informational status of sentence parts: intonation, word order and morphological markers. In restricted linguistic systems, Jackendoff (2002) suggests, information is often structured according to the FocusLast principle:

[9] A third possibility for expressing information structure is to use morphological markers. This is not possible in English, but it is used in e.g. Japanese, where *wa* is used to mark the topic of a sentence. See de Swart and de Hoop (2000).

focus information is expressed last.

2.4.2 Basic Variety

A striking property of adults learning a second language is that they are so bad at it. Children are much better than adults at copying the structures that are used in the target language; adults seem reluctant to copy sentence structures they do not understand (Klein, 2001). Especially when adults learn a second language outside the classroom (spontaneous second language acquisition), not all learners reach the state of speaking the target language flawlessly. They not only tend to have a heavy accent, but more importantly (at least in the context of this thesis), they do not apply the grammatical rules of the language they are learning correctly. Many people, especially when they receive no instruction on the language they are learning, 'fossilise' during the acquisition process (their learning does not progress any further). In the stage in which most learners fossilise, they organize their utterances according to rules that are neither part of their source language, nor of their target language.

Perdue (1993) presents a large set of longitudinal acquisition data of untutored adults learning various second languages (English, German, Dutch, French, Swedish). Source languages of the speakers varied as well (Punjabi, Italian, Turkish, Arabic, Spanish, Finnish). Based on these data, Klein and Perdue (1997) discern a language form they call the Basic Variety (henceforth BV). The BV is a learning stage during the acquisition process, during which a target-language-like lexicon is combined with semantic and pragmatic rules that are independent of source and target language. Around one third of spontaneous L2 learners fossilise in the BV stage.

In the BV, speakers have limited knowledge of the target language lexicon, but they do already use verbs, and their utterances are usually structured around the verb. They do not, however, inflect verbs. Or, if verbs get inflected, they are not inflected in a consistent way.

As we have seen above, examples of pragmatic and semantic rules governing the organisation of the verb and its complements are 'AgentFirst' and 'FocusLast'. The former, AgentFirst, is a semantic rule and tells us that the NP-referent with the highest control should come first: for most verbs, it is possible to rank each argument such that one exerts greater control than the other. For example, in a verb like 'push', the *pusher* is clearly in greater control than the *pushee*, but also in a verb like 'meet', an asymmetry of control can still be observed. In copular constructions, on the other hand, there is no such natural difference in control.

(6) le fille reste avec l'autre dame
the girl stays with the other lady
Starren (2001)

The other principle, FocusLast, has to do with the distribution of old infor-

mation (topic-information) and new information (focus-information) over an utterance. FocusLast tells a speaker to put the new information, that which is in focus, at the end of an utterance.

This pattern becomes clear in data from a film retelling task. Learners were asked to retell part of an edited Charlie Chaplin film. From a Spanish learner of French, the following two utterances were recorded (Klein and Perdue, 1997, p. 316):

(7) il [setruv] avec la fille
he (=Chaplin) finds himself with a girl

(8) il [setruv] avec Chaplin
'he' (=the girl) finds herself with Chaplin

Sentence (7) was uttered in a situation where *the girl* was focus information, whereas sentence (8) was uttered in a situation where *Chaplin* was focus information.

If a sentence is seen as consisting of two parts, the FocusLast principle could just as well have been called TopicFirst (Hurford, 2011, section 5.7). The latter term, however, suggests that there is a source for possible conflict. When the *Agent* is not the same as the *Topic*, two items compete for initial position in the utterance. This is illustrated in the situation where a Punjabi learner of English uses the following utterance (Klein and Perdue, 1997, p. 330):

(9) stealing bread girl

This utterance is used in a setting where 'girl' is clearly Focus information, such as after the question 'who stole the bread?'. An English speaker might have used stress ('THE GIRL stole the bread') to signal Focus information in this setting, or a construction like 'It was the girl who stole the bread.' The Punjabi learner of English, however, puts 'the girl' in final position to signal Focus-hood, thereby violating the AgentFirst principle.

A different speaker, an Italian learner of German, was put in a similar situation and solved the conflict differently, by saying (Klein and Perdue, 1997, p. 330):

(10) mädchen nehme brot
girl take bread

This learner follows the AgentFirst constraint, and thereby fails to convey the right Focus information.

All in all, it has been shown that adult speakers who learn a second language without receiving instruction do not simply speak an imperfect version of their target language. Nor do they just concatenate the words they have in their inventory into strings that have no further structure and function as so-called 'semantic soup'.[10] Instead, learners organise their utterances according

[10]The term 'semantic soup' is from (Anderson, 2004), where the term is applied to ape

to principles that are based on the semantic and pragmatic properties of the information that they want to convey. Although it has its limits (as we have seen in the examples with the conflicting principles), the BV is generally a quite successful means of communication.

Second language learners generally live in a community that does not speak their language, and because they have only started learning some features of the target language, they have to improvise. This situation gives rise to the properties described above. Situations that are similar to that of adult second language learners are situations in which *pidgins and creoles* emerge. These will be discussed next.

2.4.3 Pidgins and Creoles

Pidgins and creoles are language forms that emerge in situations where groups of speakers that do not share knowledge of a common language are forced to communicate with each other. Both pidgins and creoles have been described as phenomena that can tell us something about the emergence and evolution of language (Bickerton, 1990; Givon, 1997). There is little agreement, however, firstly about how to draw the dividing line between pidgins and creoles, and secondly, about the facts on the basis of which evolutionary conclusions should be drawn from these two phenomena (Botha, 2006b). In a very brief overview, I will show the different views on pidgins and creoles, and the evolutionary claims that have been made on the basis of them.

Bickerton (1981) sketches a picture of pidgins and creoles that is both very influential and controversial. Whereas it is generally accepted that pidgins are 'simpler' than creoles, Bickerton describes pidgins as crucially and categorically different from creoles in the sense that pidgins are not 'true languages' whereas creoles are. In pidgins, there is concatenation of words, but word order is variant and not related to grammatical function. Creoles, according to Bickerton, are the languages that are created by children growing up in an environment where pidgin is spoken. Despite the fact that this input language has little consistency in structure, and no grammatical features, the language that children start using is much richer in structure, and has many basic features of established human languages. This resulting language is a creole. In other words, creoles emerge rather abruptly in pidgin speaking communities and are created by children. At least, this is the case for creoles that arose out of a prior pidgin, in a population where there was enough diversity in the native languages of the pidgin speaking parents (Bickerton, 1981).

The creoles that fall within Bickerton's definition show structural similarities and these properties represent, according to Bickerton, the expression of a biological characteristic of humans: 'the capacity to recreate language in the

communication: 'the chimpanzee has a lot going on in his head at a particular moment. Some of these thoughts correspond to signs he knows, and he produces the corresponding gestures. The signs that emerge reflect his ideas, but with no particular organization apart from general contextual salience' (p. 278).

absence of any specific model from which the properties of language could be 'learned' in the way we normally learn things' (Bickerton, 1990, p. 171). This position is known as the Bioprogram Hypothesis, which says that the structure of creoles reflects our innate grammatical dispositions.

As described above in section 2.3.3, Bickerton proposes that human language emerged in two stages: the protolanguage stage, followed by a rapid transition into full language. Protolanguage did not have structurally organising principles and can be described as 'language minus syntax'. Traces of this protolanguage stage, Bickerton argues, come to the surface in, among other phenomena, pidgin languages.

As mentioned above, Bickerton's work on pidgins and creoles has been quite influential, but many of his claims have been subject to criticism. Let us start with his primary focus on creoles that originated from pidgins. In Mufwene (2007), it was shown that certainly not all creoles originated in this way. Also his claim that a creole can emerge so quickly has been disputed (McWhorter, 1997; Mufwene, 2007). Further, the claim that structures in creole languages reflect innate dispositions for grammar is challenged by creolists pointing to the fact that many structures observed in Bickerton's creoles can be traced back to their superstrate or substrate languages (Lefebvre, 2009; Mufwene, 2007).

Lastly, Bickerton's claim that pidgin languages have no structurally organising principles is disputed: various sources point out that, similarly to the Basic Variety, (early) pidgin shows evidence for the semantic principles AgentFirst and FocusLast[11] (Bickerton, 1981; Jackendoff, 2002; Roberge, 2009). Givon adds to these characteristics the phenomenon that units of information that are related to each other, are placed close to each other in an utterance (Givon, 1997).[12]

The overall picture that emerges from this discussion of pidgins and creoles is that Bickerton's account of sudden emergence of creole out of pidgin is losing ground, and there is a fair body of literature that supports the view that semantic or pragmatic principles govern utterances in (early) pidgin. However, there remain many different views on the nature of pidgins and creoles,[13] and this makes it difficult to state general claims about principles that play a role in pidgins and creoles.

The two restricted linguistic systems we have seen so far are both *spoken* systems. But also in the manual modality, interesting linguistic phenomena have been observed. These phenomena appear, again, in situations where communication via an existing conventional system is not possible and speakers have to improvise in order to make themselves understood. Next, we will focus on newly emerging sign systems. Even though these systems are signed, and not spoken, some similarities between them and the previously discussed pidgin

[11] Though 'FocusLast' is sometimes formulated as 'Topic First'.

[12] This is similar to what Jackendoff (2002) calls 'Grouping'.

[13] For instance, Roberge (2009) points out that pidgins should be seen as dynamic systems (systems undergoing change), while many other scholars (implicitly) assume that pidgins are stable systems.

and Basic Variety can be observed.

2.4.4 Emerging sign systems: Homesign

When deaf children are born into hearing families, the hearing family members do not always know any conventional sign languages. When the child itself is also deprived from contact with conventional sign languages, it will start to develop a gestural communicative system in order to communicate with its family. The sign systems that emerge in these situations are called *Homesign* and the properties of various homesign systems have been observed and analysed. Children who have developed a homesign system all develop a lexicon of stable, often highly iconic gestural signs or pointing gestures. With their system, they are able to talk about the things that are present in their direct environment, but eventually also about objects and events that are not here and now, although they start referring to these *displaced* entities later than normally hearing children (Morford and Goldin-Meadow, 1997). Quite possibly, the developmental pattern of homesign systems, which differs from a normal first language acquisition pattern in the relation between form and function, is responsible for this. In first language acquisition, children often use grammatical forms before they fully understand the semantic function they serve; they simply copy patterns they hear and start figuring out the exact meaning of these patterns only later (Benazzo, 2009). In homesign, on the other hand, the lack of a conventional system forces the children to first fully understand a given conceptual function, and then feel the need to communicate about it, before they can actually incorporate it into their sign system (Benazzo, 2009, p. 35).

In Goldin-Meadow *et al.* (2009), it is shown that there are similarities in the patterning of lexical elements in homesign systems that emerged in different countries: children growing up in Chinese, English and Turkish speaking communities all show a tendency, when they talk about a transitive action, to omit the gesture that indicates the Actor (that would be the subject in a sentence describing the situation). Moreover, they seem to have a preference to gesture the Patient (direct object) before the Act.

Thus, there are indications that semantic properties play a role in the organisation of utterances in Homesign, even though individual systems are quite different from each other in many respects. Jackendoff's principles AgentFirst and FocusLast are not reported directly in homesign, but, as mentioned, the Agent is often omitted, and the principle FocusLast is not described as such, possibly because homesign was not studied with a topic/focus framework in mind. To conclude, the fact that at least some systematicity can be found in the ordering of the constituents in homesign utterances, shows that even in a situation where a child has had no input from a conventional language, the organisation of utterances is used to convey information.

Next, I will discuss two recently emerged sign systems that are comparable to homesign because they were developed by deaf people that did not previously

have a conventional system for communication, but that have the advantage that they emerged, not in one individual, but in a *community*.

2.4.5 Emerging sign systems: NSL and ABSL

In Nicaragua, a school for deaf children was founded in 1977. Prior to the existence of the school, deaf children were mostly kept at home, and had no contact with other deaf individuals. When the school was opened, pupils were taught in spoken Spanish, which did not seem very successful, but among themselves, the children in the school started to communicate with each other using gestures and over the years, a sign system emerged: **Nicaraguan Sign Language (NSL)** (Senghas *et al.*, 2004).

Numerous aspects of NSL have been studied in different generations of signers. Senghas *et al.* (2002) showed videos of dynamic events, such as a cartoon of something rolling downward, to signers of NSL. It was analysed how signers from different generations signed the manner of movement (e.g., rolling) and the path of movement (e.g., downward), and it was found that first cohort signers expressed manner and path of movement simultaneously, i.e., using one gesture to signify both manner and path. Later cohorts used separate signs for manner and path, thus gesturing the information sequentially. This shows that, even though the manual modality has the possibility to convey information holistically, sequencing of information is preferred when a system becomes increasingly conventional.

Al Sayyid Bedouin Sign Language (ABSL) is a conventional sign system that emerged over the past 70 years in the Al Sayyid Bedouin group (Sandler *et al.*, 2005). This relatively isolated group, living in the south of Israel, has a high percentage of congenital deafness; deaf individuals are fully integrated into the social structure of the community and ABSL is generally seen as a second language of the village. ABSL is different from Israeli Sign Language (ISL); ABSL is unintelligible to ISL signers.

Similarly to homesign systems, subjects are sometimes omitted from gesture strings in ABSL. Moreover, action words (verbs) generally occur in final position. To be more precise, the dominant word order of ABSL is SOV (subject-object-verb), even though the surrounding languages of the village all have different prevalent word orders. Sandler *et al.* (2005) report that spoken Arabic dialect has SVO order, Classical Arabic has VSO order, Hebrew SVO (or V-initial) and Israeli Sign Language mainly SVO and OSV.

Having SOV order, ABSL shows clear signs of the AgentFirst principle. Another interesting pattern was found by Padden *et al.* (2010). They report that when situations are described where the agent and the patient are both human (as in the sentence 'the woman feeds the child'), signers split their utterance into two separate utterances, the first of which describes the actor and the second the patient (such as WOMAN SIT; GIRL FEED). A similar pattern, 'SVOV' ordering for events where the actor and the patient are both human, was found in Nicaraguan Sign Language (Senghas *et al.*, 1997).

Of both NSL and ABSL, different generations of signers have been recorded, and it was observed that the sign systems were simpler for early generations than for later generations. As the emerging sign systems become more sophisticated over time, the signing becomes faster and the sentences that are uttered become longer (Sandler *et al.*, 2005; Senghas *et al.*, 2002).

Now we have seen a very brief overview of different systems that arose in situations where 'normal' language could not be used. As we have seen, the systems emerged under quite different circumstances, but still, have some properties in common. What about these properties? And how can we draw evolutionary conclusions from them? These issues will be discussed in the next section.

2.5 Restricted linguistic systems: analysis and evaluation

2.5.1 Organising principles

The picture that emerges from the restricted linguistic systems described above is that situations where normal communication is impossible, as unfortunate as these might be, are a valuable source of linguistic data. The urge to communicate is apparently strong in humans, and the communication systems that emerge in situations like the ones described above can tell us something about which are the core elements of language: those that survive when most of language breaks down.

It is clear from all these systems that they are indeed restricted, in the sense that there is little or no inflectional morphology in them. It is not the case, however, that in these systems, words are just thrown together without structure or organisation. Utterances in restricted linguistic systems are organised, at least to some extent (some systems also show some variability within and among speakers). The principles according to which the systems are organised could very well be rooted in semantics. Let us have a look at Jackendoff's hypothesis that the principles AgentFirst and FocusLast are organising principles.

One thing that emerges quite clearly from the data is Jackendoff's principle 'AgentFirst': we have seen that in all systems except homesign (where the agent is often omitted from the utterance), the constituent that has the semantic role of Agent (the one that has the highest control) is put in first position.

Judging whether FocusLast can be observed systematically in the systems discussed is somewhat harder, however. First of all, not all systems have been analysed and described in terms compatible with a principle such as FocusLast. This principle is about 'focus information', which is often described as 'newness of information', and it is not always easy to see which information in an utterance is new and which is not. Moreover, it should be taken into account that the orderings found in the different systems are not always entirely independent of influences from outside (i.e., the circumstances in which they

arise). For instance, pidgin languages might be influenced by their superstrate and substrate languages. In order to verify that FocusLast is observed across restricted linguistic systems, careful analysis of data is necessary; preferably not isolated sentences, but sequences of sentences.

If we want to establish restricted linguistic systems as a source of indirect evidence about the emergence of language, there are two further domains that need additional attention. The first concerns the step from the data to the evolutionary conclusions: how do we justify this step? The second concerns the way in which data is obtained from the different systems: because all of the systems discussed here emerge in unfortunate situations, gathering empirical data is not straightforward. Both issues will be discussed below.

Subsequently, I will discuss how to proceed, i.e. how to make sure that restricted linguistic systems can be put to further use to investigate the emergence of semantic structure in human language: by 'creating' a restricted linguistic system in the laboratory.

2.5.2 Linguistic data, evolutionary conclusions and bridge theories

The linguistic data presented above show that language can exist in situations where it is hard to communicate, and that this language form is relatively successful for communication, and shows similarities across various situations. But how do we get from the observation that the data are interesting to conclusions about the emergence and evolution of language? And how can we be sure that these conclusions are justified?

In a series of publications, Rudolf Botha aims to create a framework for the analysis of restricted linguistic systems in an evolutionary context. In fact, the framework is set up not only to look at restricted linguistic systems, but also at other evidence, such as, e.g., archaeological findings. In the context of this dissertation, we will focus on restricted linguistic systems only. Botha formulates an approach to language evolution called *windows approach*. A window on language evolution is characterised as follows:

> [W]indows are shown to be conceptual constructs used for making inferences about aspects of language evolution from data or assumptions about properties of phenomena other than language evolution.

The particular kind of window that is generated by looking at data from RLS's is the *analogue window*.[14] In the analogue window, it is believed that aspects of the data observed are analogous to particular aspects in language evolution. Analogue windows are used to draw conclusions about 'internal aspects' of language evolution: they say something about the function and struc-

[14] In Botha's work, this kind of window is contrasted with the *correlate window* and the *abduction window*, but these fall outside the scope of this thesis.

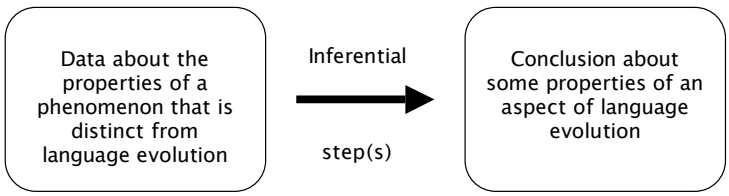


Figure 2.2: Basic structure of a window on language evolution (Botha, 2009b).

ture of early forms of language. To make this specific, the structure found in, e.g., the Basic Variety is analogue to the structure of some protolanguage. At first sight, this does not sound very different from the proposal made by Jackendoff, but Botha proposes that windows be treated systematically to make sure any inferences are made in the correct way. He describes 'window work' as a three step process, which starts with observations of some phenomenon distinct from language evolution, and results via inferential steps in conclusions about aspects of language evolution, see figure 2.2.

Each step in the process must meet certain conditions in order for a window to have merit: (a) the data on the basis of which evolutionary conclusions are to be drawn need to be accurate, (b) the evolutionary conclusions must be underpinned by a theory of linguistic entities and a general theory of what evolution involves, and (c) the inferential step must be underpinned by a 'non-ad hoc bridge theory' (Botha, 2009b). Botha has evaluated various windows on language evolution, by examining if the criteria he formulated are met for each of these windows. From a series of publications (Botha, 2006b, 2007, 2009a), the general picture emerges that it is quite hard for any window to meet all the criteria.

The current situation for restricted linguistic systems is best summarised as follows: they can potentially make a valuable contribution to the language evolution debate, but some attention needs to be paid to the underpinnings of evolutionary conclusions based on the data. Let us have a look at the conditions formulated by Botha in more detail. Concerning condition (a), it might be hard for some of the restricted linguistic systems discussed above to prove that the data are accurate (about pidgins and creoles, for example, there is a lot of disagreement on the nature of the data to begin with). I will discuss this issue in further detail below (in section 2.5.3), and subsequently present my own way to gather data from restricted linguistic systems, in a laboratory setting (in section 2.5.4).

Concerning condition (b), I have sketched a picture of the language evolution debate, and I have identified the specific topic I am interested in: the emergence of structured meaning. Moreover, I have specified the place that this topic takes in the debate. What is still lacking, though, is a characterisation of

meaning in modern language. This will be taken up in the next chapter.

Condition (c) demands that there is something like a bridge theory from the data to the evolutionary conclusions. Botha (2006a) formulates the following question, that needs to be answered for restricted linguistic systems (which are addressed here as 'degraded forms of language') in order to have a proper bridge theory:

> Why does the fact—if fact it is—that AgentFirst and FocusLast occur in some 'degraded' forms of language give these order principles their so-called evolutionary primitive character, i.e., their language fossil status? (Botha, 2006a)

Jackendoff answers this question only by indicating that the fact that AgentFirst and FocusLast survive processes in which language is degraded makes them special, but does not specify the nature of these principles any further. In fact, when discussing the step from empirical data from simple language systems to evolutionary conclusions, he adds the following caveat:

> It is of course never clear how relevant such evidence is for evolutionary concerns—in particular, to what degree their ontogeny really recapitulates phylogeny. Nevertheless, this is all the evidence we've got, so we must make the most of it, while recognizing that it should be taken with a grain of salt. (Jackendoff, 2002, p. 237)

In a sense, Jackendoff is right: making *any* claim about the emergence and evolution of language is problematic, because we cannot go back in time and check the facts. But this should not prevent us from attempting to formulate precise hypotheses and make the best use of the (indirect) evidence we have.

Fortunately, in recent publications, initial steps have been taken to formulate a bridge theory for data from the Basic Variety and from pidgin languages. These will be discussed in chapter 4.

2.5.3 Collecting restricted linguistic data

It has already been mentioned that collecting data for restricted linguistic systems is not easy. Not only do many of these systems emerge in unfortunate situations, it is also quite a challenge to get a reliable set of data. To name a few problems:

- The Basic Variety data comes from a fairly large database (Perdue, 1993), for which learners of different languages were studied in longitudinal studies. As mentioned above, the source languages that were studied were Punjabi, Italian, Turkish, Arabic, Spanish and Finnish and the target languages were English, German, Dutch, French and Swedish. Although the source languages are quite diverse, the target languages are relatively similar, and, for example, all have SVO as the dominant word order. And

this word order might have had an influence on the word order in the Basic Variety (Hurford, 2011).

- For data on pidgins and creoles, there is little consensus about the nature of the data. Not all linguists have the same conception of what pidgins and creoles are, and it is not easy to draw evolutionary conclusions from a phenomenon that is poorly demarcated (Botha, 2006b).
- Gathering data about homesign is quite labor-intensive and existing publications are often based on data from only a handful of signers.
- There is only a limited amount of data about NSL and ABSL. Moreover, these systems have developed and changed quite rapidly, which is interesting in itself, but it makes analysis and comparison to other restricted systems complex.

There is also a more general problem for restricted linguistic systems, which is the lack of control in the collection of data. Data for the systems is usually collected in a conversation-like setting. For example, playing children interacting with their peers are recorded to obtain homesign data, and interviews with language learners are held to obtain Basic Variety data. If one is interested in a certain construction, one can try to steer informants in a certain direction by asking them specific questions. This is done, for example, in the ESF project (on the basis of which the Basic Variety was described). People watched scenes from a film, and were asked to retell the story (Perdue, 1993). Still, a natural conversation setting does not lend itself to collection of very specific and controlled data.

Another general problem is that in these systems, the main focus is on language production, and no attention is paid to *comprehension* of restricted language. However, we need insights about comprehension to analyse the semantics of utterances from restricted linguistic systems.

2.5.4 A restricted linguistic system from the lab

We have observed that restricted linguistic systems are interesting and intriguing on the one hand, but they still face some problems. Problems concerning data have been formulated in the previous section, and given these problems, one could conclude that it is necessary to go out and collect more data on homesign, pidgin, the Basic Variety, or newly emerging sign languages. And I definitely think that this would help a lot, especially when similar phenomena are studied in different systems and then compared to each other. This could make a valuable contribution to the field.

But I want something different: I would like to be able to collect data on restricted linguistic systems in a very controlled manner, and study exactly those issues I am interested in. Moreover, I want to be able to study the interpretation of restricted language as well as the production. To get this, I would need a laboratory environment, where participants can be given carefully composed and controlled sets of stimuli. Fortunately, a way to 'produce' restricted

linguistic systems in a laboratory has been designed already, although it is not described in such terms.

Goldin-Meadow *et al.* (2008) describe an experiment where participants are asked to communicate about simple events presented in short animations, using only gesture and no speech. Participants are thus asked to communicate in a way they have never done before and cannot use their native language. Goldin-Meadow *et al.* (2008) show that participants from four different language groups show consistency in the ordering of their gestures, and thus organise their gestured utterances independently of their native language. In chapter 4, I will elaborate on this method and argue in further detail why the gesture strings produced by participants in this way can be seen as a restricted linguistic system. Moreover, I will show that this method allows us to look at, not only the production of improvised utterances, but also their interpretation, and I will describe the particular topics that will be investigated in chapter 5 and 6.

2.6 Conclusion

Throughout this chapter, I have shown that there are different ways to obtain indirect evidence about the evolution of language. One of the approaches presented, the one that studies restricted linguistic systems, offers a way to look at those aspects of language evolution I am interested in: the role of semantic principles in protolanguage.

We have seen data from restricted linguistic systems, and reflected on the conclusions drawn from them. It was shown that it might indeed be the case that restricted linguistic systems are governed by semantic and pragmatic principles like Jackendoff's AgentFirst and FocusLast. Some problems still remain, however, and if we want to establish restricted systems as serious (indirect) evidence about language evolution, these have to be resolved. The issues that have come across have to do with, first of all, the principles AgentFirst and FocusLast themselves. It is not always easy to determine whether these principles are at work or not, on the basis of the data, and this should be studied further. The second problem has to do with the evolutionary conclusions drawn on the basis of restricted data. The step from the linguistic data to the evolutionary conclusions needs to be underpinned by a bridge theory that tells us exactly what the nature is of the analogy drawn between restricted linguistic systems and evolutionarily early language. Thirdly, the data from restricted linguistic systems reviewed in this chapter only concerns language production; given the circumstances in which these systems emerge, it is hard to study the interpretation of restricted language. Lastly, I formulated some practical problems with the collection of data from restricted linguistic systems. In general, because restricted systems emerge in exceptional situations, data is not easily available.

To solve the last two problems, a new method to collect restricted linguistic data was announced. This will be described in further detail in chapter 4. To solve the other problems, we need to establish a firmer theoretical basis on

which evolutionary hypotheses can be built. This issue will be addressed in chapter 4 as well. But before we will be able to do that, we need to know more about what we will be studying: meaning. What is meaning? Which concepts are important when we describe meaning in natural language? What do these concepts presuppose about the emergence of meaning? These issues will be addressed in the next chapter, and this will help us to prepare the ground for restricted linguistic systems to make a serious contribution to the language evolution debate.

CHAPTER 3

Meaning and evolution[1]

3.1 Introduction

What is meaning? The answer to this question depends on who you ask. Among those who think about the nature of meaning are philosophers and linguists, and the discipline that studies meaning in natural language is generally called semantics.

> If phonetics is where linguistics makes contact with physics, and syntax is where linguistics meets mathematics and computer science, semantics is the branch of language study that consistently rubs shoulders with philosophy. This is because the study of meaning raises a host of deep problems that are the traditional stomping grounds for philosophers. (Fitch, 2010, p. 119)

Thus, in this chapter, we will look at characterisations of meaning by philosophers and by linguists. In the first part of this chapter (section 3.2) I will provide a brief overview of theories of meaning in the philosophy of language. Many — very different — proposals have been made about the nature of meaning, either in a formal, logical context, or in an informal, philosophical context. I will sketch the differences between the major existing views on meaning. It will become clear that there are three different intuitions behind these views: that meaning should be defined in terms of reference, in terms of

[1] Part of the research presented in section 5 of this chapter was originally published as Schouwstra (2008).

intentions, or in terms of conventions. These three intuitions all capture important aspects of meaning, and I will show that these intuitions play a role when meaning is studied in an evolutionary context.

Philosophers of language focus on the meaning of modern language, and do not primarily concern themselves with the evolutionary history of natural language meaning, and, as will be shown, it is not clear for every account of meaning how it would fit in an evolutionary scenario. What happens when meaning is studied in an evolutionary context? This will be the topic of the second part of the chapter (sections 3.3–3.7). I will sketch a picture of the emergence of meaning as provided by Gärdenfors (2004b). Gärdenfors draws a close connection between meaning and cognition and sketches how the emergence of sophisticated cognition was important for the emergence of modern language, and this is in fact an evolutionary version of meaning defined in terms of intention.

I will show that this assumption (roughly, 'cognition before language') is present in the work of many others in the language evolution debate. But there are also arguments against it. After discussing these arguments, I will sketch an alternative view on the emergence of language, based on the work of Ruth Millikan (e.g., Millikan (2005)). In this view, the fact that language is a public phenomenon deserves more attention; Millikan's view can be seen as an evolutionary version of conventional meaning. Thus, I will characterise two evolutionary scenarios: a cognitive/intentional and a communicative/conventional road to meaning in natural language.

The two possible roads to language are, I will claim, not mutually exclusive: they present extremes of a scale, and the actual evolutionary history of language will probably contain elements of both. The two roads, however, represent ways of thinking about language evolution, and they will prove to be helpful in the next chapter, where I will formulate a full justification for evolutionary conclusions on the basis of restricted linguistic systems.

Thus, focusing on definitions of meaning will provide tools for thinking and talking about the emergence of language. But we do not *only* need general and broad ideas about meaning. We also need to zoom in on language and meaning, to study utterances made in natural language, the meaningful units that can be distinguished in them, and the contribution these units make to the meaning of complex wholes. In the last part of this chapter (section 3.8), I will look at meaning from the perspective of a linguist. Whereas philosophers tend to prefer to sketch big pictures and philosophers of language look at language and meaning as a general phenomenon, linguists study language more closely by looking at the meanings of individual utterances. I will have a look at what linguists say about how structured meanings of utterances are built, and I will discuss two phenomena: *compositionality* and *thematic relations*. This will provide us with terms and concepts necessary for thinking further about the role of meaning in protolanguage.

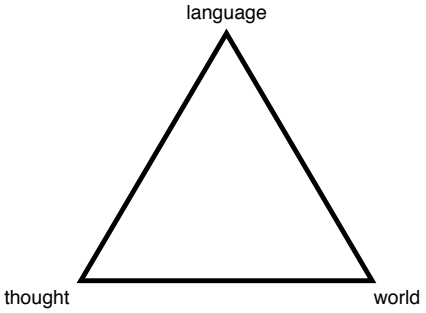


Figure 3.1: Language, thought and world

3.2 Meaning in philosophy of language

Before I discuss how meaning is defined in philosophical literature, let us have a look at some basic intuitions about meaning. When we formulate sentences, the expressions we use *refer to* things in the world. In other words, if I say: 'My bicycle is green,' I say something about a particular object, namely my bicycle, and I state that it is green. But expressions also have much to do with thought: the sentence 'My bicycle is green' has some relation with the mental representation I have of my bicycle and its properties, and whoever hears the sentence, will form a mental representation of my bicycle too.

This very simple example shows that the meanings of the expressions in our language have a certain relationship with thought, but also with the world. This basic intuition, that meaning has something to do with the relationship between language and the world on the one hand, and with the relation between language and thought on the other (see figure 3.1), proves to be important when we look at what has been said about meaning in philosophy and linguistics.

In this section, I will provide a very brief overview of some of the many existing theories of meaning that were put forward in philosophy of language and linguistics, starting out with a description of propositional semantic theories that were formulated in philosophy of language, followed by Chomsky's account of meaning, Grice's definition of meaning, and concluding with Lewis's conventional account of meaning.[2] I am aware that this is a very crude summary of what has happened in linguistics and philosophy in the past centuries, but the highlights I chose will provide us with useful tools to think further about the emergence of meaning in human language.

[2]The selection I made is loosely based on the overview provided in Speaks (2010).

3.2.1 Meaning as a proposition

In philosophy of language, theories of meaning have as a goal to describe the semantics of natural language, but in a very abstract sense. Theories like these will not look at detailed differences between the semantics of e.g. French versus the semantics of English. Neither do they go into the analysis of the semantics of one very particular kind of expression. Instead, they provide a general idea of what meaning is. Some accounts of meaning postulate meanings as entities that correspond to sentences in natural language, and discuss the content of these entities. The entities corresponding to natural language sentences are generally referred to as *propositions*; hence the title 'meaning as a proposition'.

Meaning in terms of reference

To look at propositional semantic theories, let us first look at a definition of meaning in terms of *reference*. A theory of meaning in terms of reference underlies all propositional theories of meaning, so it is important to understand this account and its problems first.

A definition of meaning in terms of reference pairs expressions with the contribution these expressions make to the determination of the *truth values* of sentences in which these expressions occur. The following examples illustrate this.

(1) Barack Obama is a Democrat.

The name Barack Obama in (1) refers to the person Barack Obama, and the predicate '. . . is a Democrat' refers to the set of Democrats. The sentence is true if and only if Barack Obama is a member of the set of Democrats (which is the case, so the sentence is true). By contrast, the following sentence is not true.

(2) Sarah Palin is a Democrat.

Sentence (2) is false, because the person denoted by the name Sarah Palin is not a member of the set of Democrats.

To summarise, the parts of the sentence refer to individuals, properties or relations, and the sentence as a whole combines the referents of the parts and thereby states the truth conditions for the sentence.

From the example above, the referential account of meaning seems intuitive and effective: sentence (1) is true, because of what the name 'Obama' refers to, and sentence (2) is false, because of what the name 'Palin' refers to. However, the analysis of meaning in terms of reference has major problems.

One of these problems is a problem with *intensional contexts*. Consider the following two sentences (taken from Gamut (1991)):

(3) John is looking for the supreme commander of the US armed forces.

(4) John is looking for the president of the United States of America.

The underlined expressions in (3) and (4) always denote the same person. Still, it might be the case that sentence (3) is true, while sentence (4) is false: it might be that John is looking for the supreme commander of the US armed forces, without realising that the president of the USA is always also the supreme commander of the armed forces.

Problems like this[3] point out that the meaning of an expression cannot simply be identified with its reference. Instead, it must have something extra, something that is called *propositional content*.

Propositional content

Using the term 'propositional content', we can formulate the problem described above as follows: it is possible for two expressions to have the same referent, but to have different contents. Propositional content can be seen as an *extra level* in between the expression and the reference of the expression.

Various ways have been proposed to define meaning as some entity that leads from the expression to the reference; see Speaks (2010) for an overview. For the purpose of this very brief overview, I will describe just the essentials of one of these theories, possible worlds semantics.

Possible worlds semantics defines the meaning of an expression s as a function from the circumstances of evaluation of s to the reference of s. In other words, an expression is not what it stands for in the world (the way it is defined in the referential theory of meaning), but rather a rule that tells you what an expression would stand for *if the world were a certain way*. This is nicely illustrated with an example in Speaks (2010):

> The idea is that the meaning of an expression is not what the expression stands for in the relevant circumstance, but rather a rule which tells you what the expression would stand for were the world a certain way. So, on this view, the content of an expression like 'the tallest man in the world' is not simply the man who happens to be tallest, but rather a function from ways the world might be to men—namely, that function which, for any way the world might be, returns as a referent the tallest man in that world (if there is one, and nothing otherwise).

Thus, in possible world semantics, an expression like 'the tallest man in the world' doesn't simply lead to one individual, but it gives us a recipe to arrive at the right individual: if John's height is the greatest of all men, then John is the tallest in the world, but if Sam's height is the greatest, it is Sam, etc. This has as a consequence that, according to the theory, we can understand the meaning of the expression 'the tallest man in the world' without actually knowing who the tallest man in the world is.

[3] More problems have been formulated; see Speaks (2010) for examples and references.

In possible worlds semantics (like in other propositional accounts of meaning), the analysis of the semantics of linguistic expressions takes a tight connection between expressions and reference as a starting point but introduces an extra layer, in order to solve problems where expressions have the same referent, but do not mean the same thing.

Thus, *reference* is an important concept when we talk about meaning in human language. It captures the intuition that we use words to refer to things in the world. Full blown accounts of meaning can be formulated on the basis of a tight relationship between expressions and reference, as long as some 'extra level' is postulated to solve problems like the one described above with intensional contexts. This 'extra level' inevitably has something to do with the *understanding* of meaning, and thus, a (loose) connection is made between language and thought.

The account of meaning discussed next takes an entirely different relationship to be a good basis for a theory of meaning: one between language and mental representations.

3.2.2 Chomsky on meaning

One of the critics of reference based views of meaning is Noam Chomsky. In this section, I will briefly summarise the view on meaning that Chomsky proposes instead. In fact, three basic assumptions are important in Chomsky's position:

- Meaning should not be analysed in terms of reference.
- Linguistics is a naturalistic enterprise.
- Naturalistic inquiry should be carried out on those aspects of language that are realised in the language user's brain, and not on language as a public phenomenon.

Chomsky (2000) argues that it makes no sense to associate meaning with mind-independent objects, because the things we refer to with language are far too nebulous to be part of a linguistic theory. As an example of the way in which objects can be nebulous, consider Amsterdam. Amsterdam is located in the Netherlands, but everyone would agree that it could in principle be moved to some other place. Alternatively, everything in Amsterdam could be radically altered over time (as many things have been), but Amsterdam would still be Amsterdam. Given these observations, it becomes hard to give a precise definition of the object that the word 'Amsterdam' refers to. This, according to Chomsky, is a reason not to attempt to explain meaning in terms of reference.

What, then, are the meanings of the words we use? According to Chomsky, we are born with lots of innate conceptual resources from which we construct concepts. These concepts are mental constructs that are realised, at some level of abstraction, in the human brain (and coded for in our DNA). The goal of linguistics, according to Chomsky, is to study those features of language that are represented in the brain. This is to be done in a naturalistic way, hoping

that that leads, eventually, to unification of linguistics with the 'core' natural sciences (Chomsky, 2000, p. 106).

The properties of structures of human language that are realised in the heads of individual speakers are called *I-Languages*, where 'I' stands for internal or individual. I-Languages are contrasted with *E-Languages*, which are what are generally understood as shared public languages. The latter, according to Chomsky, are mysterious entities that are not suitable for precise, naturalistic inquiry (for reasons similar to the reasons for not studying the referents of our terms: it is not clear where one public language stops and the next one begins).

As long as linguistics has not been unified with the 'core' natural sciences, we can describe I-Language in linguistic terms: as consisting of a *computational procedure* and a *lexicon*. The lexicon is a collection of items that have *features*. The computational procedure puts the items with their features together into complex arrays of features. The role of meaning in this process is altogether quite modest, and Chomsky even questions the rationale of the existence of semantics:

> It is possible that natural language has only syntax and pragmatics: it has a 'semantics' only in the sense of 'the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation, is actually put to use in a speech community'.(Chomsky, 2000, p. 132)

Summing up, symbolic objects generated by the I-Language do have semantic features, but the role of these features is quite modest and essentially subordinate to syntactic processes. Thus, what Chomsky says specifically about meaning is mainly that the role of meaning is rather small. Further, he does not mention the existence of compositional meaning: as far as meaning plays a role it is strictly at the level of the lexicon. Moreover, the emphasis on I-language has as a consequence that in a Chomskian model of language, language is strictly *internalistic* and *individualistic*. Only mental processes in individuals are deemed interesting, and communication among individuals plays no role at all.

How does Chomsky's view on language (and meaning) relate to the propositional accounts of meaning we have seen above? In figure 3.1 we have seen that, intuitively, language has some connection with both the objects that our linguistic expressions refer to, and the mental representations we have of the things we talk about. There are many ways to describe the differences between the two accounts of meaning we have just seen a description of, but let me express the differences in terms of the triangle: the main difference between propositional accounts of meaning and Chomsky's is that in the former, reference is important and in the latter, the emphasis is on the processes that take place in the language user. This difference is depicted in the triangles in figure 3.2 and 3.3.

Of course, there is a lot more to say about the differences between propositional meaning and Chomsky's account of meaning. But the distinction be-

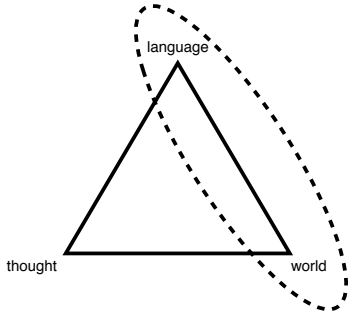


Figure 3.2: Language vs world: propositional accounts of meaning stress the relationship between language and the world, and define meaning in terms of this relation.

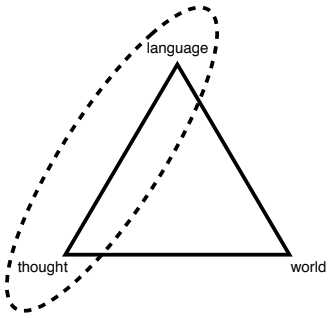


Figure 3.3: Language vs thought: according to Chomsky, the role of meaning in the study of language is rather small, but language is taken to be interesting only in an internalist sense: it is to be defined in terms of mental processes.

tween propositional meaning and 'mental' meaning seems to characterise a very essential difference between ways to think about meaning. It is therefore not surprising that this distinction can be seen in various other overviews of theories of meaning (such as the one in Fitch (2010)).

There is a second account of meaning that takes the relation between language and thought to be important: the account of meaning formulated by Paul Grice (Grice, 1969). This account of meaning differs from Chomsky's in the sense that, first of all, it gives meaning a serious and central place in the analysis of human language, and secondly, it gives a central place to the property of language that it is used for *communication*. By introducing communication, Grice introduces a dimension that is not present in the triangle in figure 3.3. The next section sketches Gricean meaning.

3.2.3 Grice: meaning and intention

The aspect of meaning that is stressed in the work of Paul Grice is the fact that a speaker of a certain utterance that bears a meaning *intends* something with this utterance. Let us have a look at the following example, from Grandy and Warner (2009):

> [I]magine you are stopped at night at an intersection, when the driver in an oncoming car flashes her lights. You reason as follows: "Why is she doing that? Oh, she must intend me to believe that

> my lights are not on. If she has that intention, it must be that my lights are not on. So, they are not."

This example shows that there is an intuitive relation between the meaning of a sign (in this case not an utterance, but the flashing of lights) and intentions and beliefs. Grice defines the meaning of utterances in a similar way. His definition of speaker meaning is the following:[4]

> a means p by uttering x if and only if a intends in uttering x that
>
> 1. his audience come to believe p,
> 2. his audience recognize this intention, and
> 3. (1) occurs on the basis of (2)

Thus, according to Grice, meanings of expressions are to be explained in terms of what speakers mean by utterances of them. Meanings of the utterances of a speaker are to be explained in terms of the intentions of the speaker.

Moreover, the intention of the speaker should, according to Grice, be recognised by the hearer. By putting the roles of speaker and hearer in such a central position in his theory of meaning, Grice not only puts an emphasis on the relationship between utterances and intentions of the speaker, but he also takes *communication* to be important in the analysis of language. Below in section 3.2.4, however, we will see an account of meaning where this is done even more radically: Lewis's conventional account of meaning. Before that, let us have a look at the differences between a Chomskian and a Gricean account of meaning.

Above, I have sketched a dichotomy between theories of meaning that emphasise the relation between language and the world (or truth values) versus theories that emphasise the relation between language and the mind. We saw that an important body of theories in philosophy of language has emphasised the relation between language and the world.

The view of meaning in Chomsky's work, as well as that in Grice's work, however, has a very different emphasis: meaning is seen as something that has much to do with what is going on in someone's mind when he uses language. Chomsky takes the study of language to be essentially internalistic, and what is important about meaning (as far as meaning plays any role at all) is defined in terms of mental processes. Grice defines meaning in terms of the *intention* of the speaker.

What are the main differences between Chomsky's account of meaning and Grice's? Let us have a look at a distinction once made by David Lewis (1970):

> I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one

[4] As formulated in (Grice, 1969, p. 159).

> of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics. (Lewis (1970), quoted in Speaks (2010))

Lewis distinguishes between accounts of meaning where the world is in a central place, and accounts in which the psychology of the speaker is more important, just like we did above. But Lewis's distinction gives rise to a second way of dividing theories. This way is described in Speaks (2010) and it takes as a dividing line the way in which different accounts of meaning explain or describe meaning. In fact, two kinds of questions can be answered when talking about meaning:

1. What is the meaning of a certain expression (for a particular person or group)?

2. In virtue of what facts about that person or group does an expression have a certain meaning?

Theories that have the second question in mind are called *foundational theories of meaning*. Their goal is not to give an account of some already existing entity, meaning, but to dig deeper: they would like to give an account of *in virtue of what* expressions mean what they mean.

The difference between propositional accounts of meaning and Chomsky's account of meaning on the one hand and Grice's account of meaning on the other can be described in exactly these terms. in both propositional accounts of meaning and in Chomsky's, meaning is taken as a given. In the former, meanings are taken to be abstract entities that should be described by some theory, and in the latter, what is taken to be important about meaning is innate.

Grice, when studying meaning, does not take meaning for granted. In other words, he is not only interested in the kinds of meanings that are expressed in natural language. By defining meanings of expressions in terms of intentions of language users, he attempts to address the facts by virtue of which expressions have a meaning.

In fact, his description of natural (e.g., 'Those spots mean measles') and non-natural meaning (e.g., 'Those three rings on the bell (of the bus) mean that the bus is full') in Grice (1975) points towards an analysis of the *origin* of meaning, although he explicitly addresses this topic only in a much later paper, Grice (1982).[5] In that paper he proposes that non-natural meaning is a descendent of natural meaning.

A fundamental difference between natural and non-natural meaning is that natural meaning is factive (the spots that mean measles cannot stop meaning measles), and non-natural meaning is non-factive (the bus driver can ring the bell three times while the bus is not full at all). As an example of something

[5]See also the reconstruction in Avramides (1997).

that is a case of natural meaning but can grow towards non-natural meaning, Grice mentions groaning. A groan is a natural sign of pain, when someone groans involuntarily. It is possible, however, to produce a groan voluntarily, to deceive the audience. Then, the perceiver will be led to the conclusion that the groaner is in pain, but this happens on the basis of the natural meaning of groaning.

Using this example, Grice shows, in a tentative way, that there is a conceptual connection between natural and non-natural meaning. This account has not been worked out further, and Grice did not intend to sketch a historical or evolutionary picture of meaning, but the fact that he addresses the transition from natural to non-natural meaning shows that Grice's interest really lies in *explaining* meaning, rather than just describing it.

Thus, Grice is interested in where meaning comes from, and does not take meaning as given, as is done in propositional theories of meaning and in Chomsky's conception of language. By taking up this attitude, Grice expresses his interest in language as a *means of communication*. In fact, communication takes a central place in Grice's definition of meaning: he defines meaning in terms of what someone intends to convey in an act of communication. This emphasis cannot be seen in Chomsky's work.[6]

Even though Grice is interested in language as a means of communication, he defines meaning entirely in terms of intentions of individual speakers. In the next section, we will focus on an account of meaning that takes the *public* character of language to be a central aspect: the conventional account of meaning formulated by David Lewis (Lewis, 1975).

3.2.4 Conventional meaning

We have now seen accounts of meaning that stress the relation between language and the world, or between language and the mind. There is a third way to look at meaning and in this third conception of meaning, the fact that language is a conventional activity is taken to be central. When I use a certain utterance to describe a state of affairs, I do this because I know that this utterance will be meaningful to the listener. How do I know that an utterance will be connected to a certain meaning? Because I know that there is an existing convention, in the language I speak, to connect that meaning to that utterance. More specifically, if I say

(5) Mijn fiets is groen.

to a speaker of Dutch, I know that this speaker will come to understand that my bicycle is green, because he knows what the utterance means.

The first philosopher to provide a conventional account of meaning was David Lewis (in Lewis (1969) and Lewis (1975)). Above we have already seen that Lewis (1970) distinguishes two topics when it comes to the characterisation

[6] To the contrary; he is not interested in language as a public phenomenon.

of language: language as a formal system, and language as a social phenomenon. Lewis (1975) uses this contrast to build a conventional account of meaning: a distinction is made between *languages* (specific instances) and *language* (the phenomenon).

Languages are sets of strings and meanings such that to each string, a meaning is assigned. Those meanings he describes in terms of possible worlds (like we have seen above in the section on propositional meaning). Thus, a language is a set of pairs of strings and meanings (which are sets of possible worlds): each string means something, and this meaning is expressed in terms of possible words.

Language, on the other hand, is mainly a social phenomenon. It is a sphere of human action in which speakers produce strings of sounds in order to bring about beliefs or action in a listener (note that this characterisation has very clear Gricean influences). Within a linguistic community, there are regularities between (1) beliefs held by a speaker and the utterances he produces, and (2) utterances of a speaker and the beliefs or actions they bring about in listeners. These regularities (or at least some of them; see (Lewis, 1975, p. 135)) are conventions of language.

Now we know that languages are sets of strings and their ascribed meanings, and language is a convention-governed social activity. The two can be connected to each other in the following way:

- A language L is *used by* a population P if and only if there prevails in P a convention of truthfulness and trust in L, sustained by an interest in communication.

In a Lewisian account of meaning, propositional meaning (meaning in terms of possible worlds) is connected to Gricean meaning (meaning in terms of a speaker's intention). Propositional accounts of meaning make meaning into something rather rigid, by connecting utterances to meanings without taking speakers into account. Gricean meaning, on the other hand, makes meaning totally dependent on the speakers, without explaining how speakers and hearers coordinate their communicative intentions by exploiting a pre-existing practice (Rescorla, 2011). Lewis fills this gap by using the concept of convention.

To wrap up, what is it for a linguistic regularity to be a convention? Speaks (2010) summarises (and simplifies) Lewis's conventional account of meaning as follows:

> A sentence S expresses the proposition p if and only if the following three conditions are satisfied: (1) speakers typically utter S only if they believe p and typically come to believe p upon hearing S, (2) members of the community believe that (1) is true, and (3) the fact that members of the community believe that (1) is true, gives them a good reason to go on acting so as to make (1) true. (Speaks, 2010)

Thus, in Lewis's account of meaning, language as a medium for communication is taken even more seriously than in Grice's account, and 'linguistic

conventions' are used to explain how speakers understand each other's utterances.

3.2.5 Conclusion

In my overview of theories of meaning I have attempted, not only to describe some of the main theories of meaning, but also to give an idea of how the theories relate to each other. We have seen that three intuitions play a role in defining language:

reference The things words and sentences refer to are important when we want to know their meaning.

intention and belief When a speaker makes an utterance, he typically intends his audience to come to believe something.

convention Our utterances have the meanings they have because these utterances have been used by other speakers in similar situations.

These three intuitions can be seen as competing accounts of meaning, but they all capture an aspect of natural language. We will see that all three concepts will play a role in the next part of the chapter, where the *emergence* of meaning in human language is discussed.

Speaking about meaning in the context of the language evolution debate poses certain demands on one's claims: an evolutionary account of meaning has to take into account the fact that there has been a transition, from animals with very primitive communicative means and limited cognitive skills, to humans with sophisticated cognitive and communicative capabilities. When embedded in the language evolution debate, we cannot simply assume that human skills are of a different kind than those of animals and then continue to describe meaning in an exclusively human fashion. Put differently, an evolutionary account of meaning cannot take meanings as given entities; it has to say something about 'where they come from'.

Formulated thus, it would seem that in the language evolution debate, accounts that define meaning in terms of reference cannot play an important role, because they define meaning for existing languages without going into the question where meaning comes from. But this is not true. Such accounts capture the intuition that we use language to refer to things in the world, and they have an appealing account of how the meaning of a sentence depends on the meanings of the parts of that sentence. In other words, propositional accounts of meaning provide a way to describe how the meanings of sentences are built up compositionally. The only thing they do not provide is a way to tell why meaning is there.

> This very propensity, the built-in assumption that "words mean things", is perhaps the most basic biological prerequisite of the

> semantic component of human language. It is this assumption, I suggest, that leads humans to attribute magical powers to names and words, and it is this same intuition that underlies the realist stance towards linguistic meaning. [...] The evolution of this referential assumption is one of our core *explananda* in a theory of the evolution of language. (Fitch, 2010, p. 125)

Thus, propositional accounts of meaning capture a central aspect of natural language: that words are used to refer to things in the world.

A Gricean account of meaning defines meaning in terms of what the speaker of an utterance intends, and thus makes language depend on the mind of the speaker. Thereby, Grice avoids making the assumption that meaning is something given. If we would put a Gricean account of meaning in an evolutionary context, we could picture the emergence of meaning in human language as a process that comes along with the emergence of the human mind. An example of such an account is given in section 3.4.

Lewis defines meaning in terms of convention. Speakers use utterances because they belong to a certain linguistic community and they know that certain utterances apply in certain situations. In this account, meaning is defined with an appeal to previous language users. An evolutionary picture of the emergence of meaning in human language in which this idea takes a central position is sketched in section 3.6.

3.3 Meaning in an evolutionary context

3.3.1 New demands on accounts of meaning

In the previous section, we have seen various theories of meaning and we have distilled three central ideas that are important when meaning is concerned: reference, belief and intention, and convention. These central ideas will continue to play a role in the following chapters, where we are going to look at empirical evidence about the emergence of meaning and language. But before we do that, we will first have a look at more recent work in philosophy, in which the nature of meaning is discussed explicitly in an evolutionary context. In other words, we will now broaden our view and look at, not just characterisations of natural language meaning, but proposals that describe how meaning in human language came about.

As stated above, when looking at accounts of the emergence of meaning, different things become important, because we can no longer take meaning 'for granted'. In other words, we cannot just assume that meaning is out there and that the concept of meaning will take care of itself. This is so, because we are no longer only looking at human language. By considering the emergence of meaning in language, it is the transition from animal-like communication systems to human language that is under consideration. The concepts that are

used to describe this process have to be suitable to describe this transition, and it is no longer enough to use concepts that are only applicable to humans. This need for clear concepts has been recognised by other authors too:

> One of the primary needs is to formulate a lasting framework of *properties and principles* that can be understood across disciplines, and that can be used as a set of standards for comparison among species. (Oller and Griebel, 2004, p. 3)

The goal in this part of the chapter is to look at such *properties and principles* that have been put forward in the literature. It is very hard, however, to develop a conceptual apparatus that will be used to describe the emergence of meaning, without actually saying something about the emergence of meaning itself. And this is exactly what we will see: concepts that were proposed in the literature come with a set of hypotheses about the nature of the process of the emergence of meaning. My task is now to present concepts proposed in recent literature and the evolutionary assumptions that come with them.

I will start out (in section 3.4) by describing the work of Peter Gärdenfors, and focus on the concepts he uses to describe animal and human communication and cognition, and the evolutionary trajectory sketched by him. It will become clear that Gärdenfors emphasises the role of cognitive structures in the evolution of language, and he argues that sophisticated cognitive structures were present in the last common ancestor of humans and chimpanzees. In other words, he claims that sophisticated semantic structures in cognition predated sophisticated structures in language.

After describing Gärdenfors's work, I will discuss the work of other authors, who seem to make similar assumptions about the relation between language and mental representations (in section 3.5.1). I will point out that these underlying assumptions play a role when empirical data is interpreted, and sketch some problems with the Gärdenfors-like line of thinking (in section 3.5.2).

Subsequently, I will give an overview of the philosophical work of Ruth Millikan about the emergence and nature of meaning (in section 3.6). Her approach to natural language meaning is slightly different from Gärdenfors's, in that she emphasises the conventional character of language. Millikan does not endorse the view that sophisticated cognitive skills predate language, and she lays a great emphasis on the way in which language has changed the way in which we think. Taking notice of her view might help us to find, eventually, a balanced starting point to do empirical work from.

As mentioned above, the accounts of meaning in this part of the chapter are not about meaning in human language alone. In fact, they describe what role meaning plays in the behaviour of animals, and then speculate about the trajectory that takes us from animal meaning to human meaning. Both Gärdenfors and Millikan refer to studies of animal behaviour in order to show how the concepts they develop apply. In the discussion of their work, I will focus on two animal species in particular. The first is genetically very much unrelated

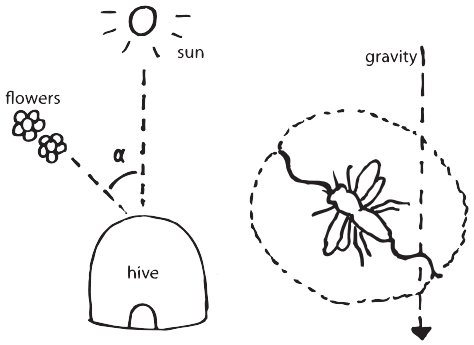


Figure 3.4: This illustrates how the angle of the waggling part of the dance corresponds to the direction of the food source: the angle between the waggling part and a vertical line is compared to the angle between the line from the hive to the sun and the line from the hive to the food source.

to humans, the honeybee. The second is genetically quite closely related to humans, the chimpanzee. Let me introduce these briefly, and explain why their behaviour is interesting in the context of this chapter.

3.3.2 Honeybees: wonders of animal behaviour?

Honeybees are an interesting species to study in the light of language evolution because they are genetically quite remote from humans on the one hand, but they have a quite sophisticated means of communication on the other: the honeybee dance. Honeybees have a very specific method of communicating about the location of food sources. Whenever they come from a food source and arrive at the hive, they make specific movements, while waggling their tails. This is what is called the honeybee dance. If the food source is relatively close (closer than approximately 150 meters) to the hive, they dance in circles, and when the food source is further away, the dance has a figure-eight shape. Different properties of the dance correspond to different properties of the food location, in the following way:

- The angle of the middle axis of the figure-eight corresponds to the direction of the food source. (See the illustration for a more detailed explanation.)
- The speed of the dance corresponds to the distance of the food source.
- The liveliness of the dance corresponds to the desirability of the food source.

And the bee dances not only for food, but also for other things, whenever they are needed, such as waxy materials, when a new hive is being built, and possible locations for a hive when a subpart of the bee colony is planning to move out of the existing hive. The latter application of the dance is especially impressive. When a new hive location is being searched for, several bees fly out to explore the environment. When they come back they make a dance about some possible hive location they have spotted. Then, the different dances of the scouting bees are compared, and the most desirable location is picked.[7]

Many researchers have been impressed by the honeybee's communicative behavior. Gould and Gould (1995), for example, decribe it as "one of the seven wonders of animal behavior". Some researchers even call the bees' communication system a language. Crist (2004) presents a list of prominent biologists, who all claim that the honeybee dance has a property that has been defined as a core property of human language: symbolism (see (Crist, 2004, p. 19)). There have also been researchers who claim that it would be a mistake to call the honeybee communication system a symbolic language (see Crist (2004) and, e.g., Sjölander (1997)).

Sjölander (1997) warns us that when we describe the behaviour of animals, using concepts that apply well to humans, we might overestimate animals:

> [A]n important lesson is that we may—in the myriad of different cognitive systems that the animal kingdom offers—find many phenomena which in humans are explained by a rather complicated cognitive mechanism, but in animals may be the effect of a direct-wired connection evolved for the specific purpose, not as an effect of a general cognitive capacity.

In this line, Sjölander argues that the honeybee dance cannot be symbolic. Rather, he describes the dance as highly 'ritualized intention movement'. Intention movement is a term from biology for an incomplete pattern of behavior that provides information to other animals.[8] According to Sjölander, the honeybee dance is nothing more than a concatenation of initial flight movements, broken off and started over again. The angle of the waggling part corresponds to the direction of the food location, simply because the initial flight movement is one *in the direction of* the food source. These intention movements can be used by other bees to predict certain things, but this is not the same as saying that the dancing bee is intentionally transmitting symbolic information.

To conclude, honeybee dance is certainly intriguing. It has led researchers to saying mystical things about bees, and caused them to make grand claims about the status of honeybee communication, even though honeybees are, genetically speaking, quite remote from humans. The honeybee language discussion shows us that it is important to have clearly defined concepts, and that we should

[7] See Gould and Gould (1995) for a very extensive description of the honeybee's communicative behavior.

[8] See also Tinbergen (1964).

be careful with our use of concepts that apply well to human behavior, in describing that of animals.

3.3.3 Chimpanzees: still waters run deep?

The chimpanzee, it is said, is the closest genetic relative of humans, which means that they share the most recent common ancestor with humans. In fact, the chimpanzee family (*pan*) consists of two subspecies: the common chimpanzee (*pan troglodytes*) and the bonobo (*pan paniscus*). The genetic similarity between humans and *pan troglodytes* is around 96 to 98.75 percent, and chimps and humans diverged from their common ancestor about 5 to 6 million years ago.[9] Thus, the divergence of humans and chimps is relatively recent, and their genetic makeup is quite similar. Still, there are very important differences in human and chimpanzee *behavior*: whereas humans have a very sophisticated means of communication (language), chimpanzees don't; they do communicate, but their 'utterances' lack the syntactic and semantic complexity of human utterances.

It has become clear from research that, even after a lot of effort is put into trying to teach them language, apes only have a rudimentary ability to build and comprehend sentences.[10] But on the other hand, apes score much better if you look at their linguistic capacities in terms of communicative goals, instead of syntactic structure.

Given the apparent discrepancy described above (little genetic difference between humans and chimps, but large behavioral, and linguistic, differences), a huge variety of opinions exists about the *cognitive* capacities of chimpanzees: how is the external world mentally represented in the chimpanzee? There is reason to believe that, despite the fact that they have no language, chimpanzees have a rich mental life (because of the genetic relatedness to humans), but on the other hand, there is reason to believe that the chimp's cognitive capacities are restricted (because of the absence of a complex language).

Even though numerous behavioral studies have been carried out, there is no conclusive evidence for either position. We will focus on this debate below, after the description of Gärdenfors's evolutionary account of meaning.

[9] About both the issue of similarity and divergence, there is quite some discussion. But for both topics, the discussions are technical, and not relevant here. See e.g. Navarro and Barton (2003) and the subsequent discussion for more details.

[10] Much disagreement exists about *what exactly* primates are capable of. See e.g. the contrast between Savage-Rumbaugh *et al.* (1993) and Wynne (2004), about the linguistic abilities of a bonobo named Kanzi.

3.4 From cognition to meaning

3.4.1 Gärdenfors: cued and detached representation

Gärdenfors sketches the emergence of complex meaning in language, and he does this by first describing the emergence of *mental representation* and subsequently connecting this to communication.

The behavior of very simple organisms is guided, to a great degree, directly by what they perceive from the external world. This very direct perception-based behavior is still present in humans, for example when we touch something very hot. But our relation with the world consists of more than only direct sensory information. We categorize what we perceive, and think about things even when they are not there. In other words, we make representations. Also it can be said of many animal species that they categorize the world, or think about things that are not present, though not as extensively as humans do. In order to describe the differences between animals and humans, Gärdenfors introduces two notions: *cued* and *detached* representations.

"A cued representation stands for something that is present in the current external situation of the representing organism." (Gärdenfors, 1996a, p. 266). Detached representations, on the other hand, "may stand for objects or events that are neither present in the current situation nor triggered by some recent situation" (Gärdenfors, 1996a, p. 266). To clarify the role of detached representations, Gärdenfors relates it to an idea from Craik (1943).

> If the organism carries a 'small scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which are the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer and more competent manner to the emergencies which face it (Craik, 1943, p. 61, quoted in Gärdenfors (1996b)).

Gärdenfors defines the 'small scale model' as the collection of all detached representations the organism has in its repertoire. In other words, all things the organism can actively think about.

Gärdenfors describes the evolution of human thinking as a process in which more and more representations become detached. The distinction between cued and detached, however, is not a sharp one: it is possible for a representation to have a *degree of detachment*, and representations with a detached flavor are present in animals. An example of this is the chimpanzee who fetches a stick to fish for termites. The behavior of the chimpanzee can only be explained if we assume that it has a detached representation of the stick (Gärdenfors, 2004a, p. 238).

On the level of utterances, Gärdenfors uses the distinction between signals and symbols that was introduced by C.S. Peirce.[11] The reference of a symbol is a detached representation; a signal refers to a cued representation (Gärdenfors, 1996b). Because a signal refers to a cued representation, and a cued representation is always about things that are present in the current outer environment, signals can be described as referring to things in the outer environment. Symbols, on the other hand, refer to detached representations, so they can be described as referring to things in the 'inner environment'. Because language is symbolic, possession of language presumes, according to Gärdenfors, something like an inner environment.

3.4.2 The path to human cognition and language

Now that we have seen the basic elements of Gärdenfors's conceptual framework, let us see how they are put together.

For his sketch of the transition from animal to human, Gärdenfors distinguishes 6 kinds of communication systems, as depicted in figure 3.5. The two dimensions along the lines of which the systems are formed are the type of representation (cued or detached) on the one hand and complexity of the communication system on the other. In communication systems, Gärdenfors distinguishes three levels of complexity. Systems with **single elements** are communication systems where only single signs, icons or symbols are used. In **compositional systems**, signs or symbols are combined in a compositional way. Systems with **grammar** are systems where the composed expressions contain different kinds of grammatical markers, and constraints on word order (Gärdenfors, 1995).

In the resulting six types, Gärdenfors envisages the following evolutionary order:

- The starting point is type 1, primitive animal signals, which are cued and have no structure.
- Then, type 4 follows, where the 'inner environment' is developed and referred to with one-word utterances.
- Then, limited combinations are introduced: a communication system of type 5, protolanguage. In the protolanguage stage, symbols are combined, but utterances do not contain grammatical elements. As an example of such communication systems, Gärdenfors quotes Bickerton (1990), where it is proposed that living fossils of once-existing protolanguages can be found in e.g. child language and first generation pidgin languages.
- Eventually, after the development of more arbitrary symbols and syntactic rules, grammaticality arises, and we arrive at type 6.

If the path to human language features type 1, 4, 5 and 6 respectively, what about type 2 and 3? No example in practice of a type 3 communication system

[11] See Atkin (2010).

	Single elements	Composition	Grammar
Cued representations	Type 1 Animal signs	Type 2 Bee's dances	Type 3 Ø
Detached representations	Type 4 One word language	Type 5 Protolanguage	Type 6 Full language

Figure 3.5: Six types of communication systems, according to (Gärdenfors, 1995, p. 7).

has been found, Gärdenfors claims. And communication systems of type 2 are rare, Gärdenfors points out, but they are the ones that are found in honeybees.

Gärdenfors's analysis of the honeybee dance is perfectly in line with the position we have seen above: that we should be careful in applying the term 'symbolic behavior' to the bee's dance. So Gärdenfors does not call honeybee dance a language. But how does honeybee behavior fit into Gärdenfors's framework? In order to answer this question, let us focus on two supposed properties of the honeybee dance that Gärdenfors comments on in his work.

First of all, the honeybee dance, supposedly, displays **displacement**, in the sense described in Hockett (1960): by dancing, the bees transmit information about things that are not here and not now. In Gärdenfors's terms this would translate into the bees having detached representations, and communicating about them. But Gärdenfors does not share this view. Referring to the fact that honeybees only dance right after returning from a food source, he claims that honeybees use their dance 'in a cued manner' (Gärdenfors, 1996b).

A second property of honeybee dance that comes across in Gärdenfors's work is **compositionality**. Bees convey information about the direction and the distance of the food source in one dance movement, in such a way that one aspect of the dance corresponds in a systematic way to one aspect of the information that is given, and another aspect of the dance to another part of the information. Hence, the honeybee dance is described as being compositional. Gärdenfors agrees with this view, and the concept of compositionality occupies a special position in his picture of the transition from animal to human.

Gärdenfors's schema is not very fine-grained, and it might benefit from further specification. But with the schema, he specifies his hypothesis about the emergence of language in the course of evolution. Moreover, along the way, he shows that the concepts of cued and detached representations are useful for

sketching a picture of the evolution of language. Much of his position, however, remains unarticulated. In order to get a clearer and more explicit picture of Gärdenfors's position, I focus on two properties of the scheme and the way in which the concepts 'cued' and 'detached' are put to use in it. First of all, in figure 3.5, two dimensions are discerned: one about mental representation (the vertical dimension) and one about (complexity of) communication (the horizontal dimension). This is not without reason: Gärdenfors is of the opinion that the dimensions of speech and thought are independent, and should be analyzed independently. Secondly, when we focus on the evolutionary order of the different communication systems Gärdenfors proposes, the step from cued to detached representations was made first, and the complexity of communication increased only after that.

These two aspects of Gärdenfors's view become important when it comes to the description of chimpanzees in terms of the framework. Speech and thought should be analyzed separately, but there is a certain dependency between the two: one needs to have an inner environment in order to have symbolic communication. In other words, mental capacities are prior to communicative capacities. In Gärdenfors (2004b) a quite specific list of mental capacities is given that are prerequisites for language, but it is possible, as Gärdenfors posits, to have many cognitive functions without actually having language (Gärdenfors, 2004b, p. 171).

Gärdenfors's way of describing things is compatible with a picture of chimpanzees (and other primates) that have a relatively rich inner life, but no advanced communication system. And this fact will become important below, when we compare this view to that of other authors. But before we do that, there is one last facet of Gärdenfors's view I would like to say something about: how does Gärdenfors describe the goings on in human minds and human language (i.e., modern language)?

3.4.3 Gärdenfors on meaning in modern language

We have seen above that Gärdenfors prefers to analyze mental representation and communication separately. When it comes to the content of communicative utterances in human language, however, he postulates a quite strong relation: the contents of utterances in human language correspond directly to a mental representation. This is directly visible in the way Gärdenfors defines symbols (see above): as referring to things in the 'inner environment'. Gärdenfors realizes that this position is not universally accepted:

> If symbols refer to the inner world, the meanings of words must be located in the head and not out in the world. This opinion stands in sharp contrast to many philosophical theories that claim that language is about the external world. (Gärdenfors, 2004b, p. 145)

I will briefly go into the specifics of the philosophical discussion that is connected to this below, but see also Gärdenfors (1993) for a full specification

of his position and an embedding in the current philosophical literature. For the moment, let me illustrate how Gärdenfors's position concerning language connects to his hypothesis about the order of things in the course of evolution. Note the following claim about symbols:

> Symbols referring to something in the inner world of a person can be used to communicate as soon as the listeners have, or are prepared to add, corresponding representations in their inner worlds. (Gärdenfors, 2004b, p. 145)

> Meaning is conceptualization in a cognitive model (not truth conditions in possible worlds). Gärdenfors (1999)

From these claims it becomes clear that Gärdenfors's position concerning this point connects well to his hypothesis that in the course of evolution, detached representations are first fully developed, and communication becomes more complex only later: for symbols to refer to the inner environment, it is necessary that the inner environment be 'ready' first. This claim is central to Gärdenfors's position.

To sum up, Gärdenfors sketches an evolutionary trajectory where mental capacities develop prior to communicative capacities. This is compatible with a picture of meaning that defines modern language meaning *in terms of* mental representations. I will call this view of the emergence of meaning *the cognitive road to meaning*. In recent literature on the emergence of language, many authors seem to have such a picture of meaning in mind. I will review some of these in the next section, and subsequently discuss criticism of the emphasis on cognition.

3.5 Evaluating the cognitive road to language

Above, we have seen that the cognitive road to meaning is a picture of the emergence of meaning that postulates that in the course of evolution, sophisticated mental representation developed first, and complex meaning in language emerged only afterwards.

3.5.1 In defense of a cognitive road to language

Jim Hurford's 2007 book 'The Origins of Meaning' offers a very rich and interdisciplinary inquiry into the emergence of meaning. It describes the cognitive capacities of many animals, quoting empirical results from behavioral biology and psychology and other areas, and proposes an evolutionary trajectory where cognitive representations of animals are the ancestors of abstract thought in humans and of meaning in human language.

Hurford shows that many features of human language have their roots in animal cognition. An example is the connection between object permanence

(to be aware that an object continues to exist, even when it is not perceived) in animals and *displaced reference* in language. When a speaker expresses information about an event in the past or the future, she refers to something that is not present 'here and now'. This property of human language, that it can be about things that are remote in time or space, is called *displaced reference* and has been indicated as interesting from the point of view of language evolution as early as the 1960s. Charles Hockett, in an article on the origin of speech (Hockett, 1960), lists four design features that are unique (or practically unique) to human communication. Displacement is one of them:

> Man is apparently almost unique in being able to talk about things that are remote in space or time (or both) from where the talking goes on. (Hockett, 1960)

Displaced reference is present in human communication, and seems to be absent in animal communication. But, according to Hockett, it would not be in principle impossible to have displacement in simple communication systems:

> It is also possible to see how faint traces of displacement might develop in a call system even in the absence of productivity, duality and thoroughgoing traditional transmission. (Hockett, 1960)

Before we look at the ways in which displaced reference is realised in restricted linguistic systems, we will briefly look at a connection between displaced reference and the cognitive capacity of *object permanence*.

The fact that displaced reference is present in human communication and not in animal communication, but that faint traces of it might develop in simple call systems, makes the notion interesting to study in the light of language evolution. Moreover, displaced reference seems to exist not only in language:

> Something akin to communicative displacement is involved in lugging a stick or a stone around; it is like talking today about what one should do tomorrow. (Hockett, 1960)

Thus, displaced reference is something that not only belongs to language but is recognised in behaviour in a more general sense. Hurford (2007) suggests that the property of language called displaced reference has its roots in the cognitive systems of the language-using individuals as the capacity of *object permanence*. Object permanence is the capacity of being aware that an object continues to exist, even when it is not perceived. The emergence of this capacity was first studied in children (Piaget, 1954), and later also in different animal species (see Hurford (2007), p. 38–40, for a brief overview and references).

When an animal has achieved object permanence, it is aware that an object continues to exist, even when no sensory information about the object is available. There are an enormous number of articles about object permanence in the literature on animal cognition. In these articles, results are presented

from tests done on various animals, in order to see whether they have achieved object permanence. To test an animal for the capacity of object permanence, two kinds of test can be distinguished.

Visible displacement: an object is visibly (to the animal subject) moved behind an obstruction, and the animal is tested to see whether it searches for the object behind the obstruction.

Invisible displacement: an object is, in sight of the subject, placed inside a box. Then the box is moved behind an obstruction and the object is removed from the box, without the animal seeing this. Then the subject is shown the empty box, and the experimenter observes whether the animal looks behind the obstruction. (Hurford, 2007, p. 38)

As might be clear from the descriptions, invisible displacement is a more complex task than visible displacement. The results of these tests, as summarized in Hurford (2007) for different animal species, are the following:

- Most vertebrates that were tested (including avians, non-primate mammals, and primates) succeed in the visible displacement test.
- Some animals, including domestic dogs, parrots and apes succeed in the invisible displacement task.
- Every animal that is able to do invisible displacement is also able to do visible displacement.
- Within primates, the closer the genetic relation to humans, the greater the chance that the animal will succeed in the invisible displacement task.
- Surprisingly, parrots perform better than some monkeys.

Thus, there are animals that solve the invisible displacement task, and this shows that in these animals, the representation of an object is controlled by knowledge (or memory), and not merely perception. Moreover, the fact that primates that are closest related to humans perform best at the displacement task suggests that displacement is a key capacity in the development of animals towards a linguistic capacity. Or, to put it in Hurfords words:

> The capacity to know something about an object, even when 'it isn't there' is a first step along the road to the impressive characteristics of human languages, their capacity for displaced reference. (Hurford, 2007, p. 72)

Here a path is sketched, starting from animal cognition, that results in displaced reference in human language, on the basis of evidence from comparative studies.

Another example that Hurford provides is the relation between *transitive inference*, and *social 'reasoning'*. Transitive inference is the human ability to conclude, given that A is taller than B and B is taller than C, that A is taller than C (given that 'taller than' is a transitive relation). Social 'reasoning' is observed in social animals like primates: individuals are often able to keep track of e.g. dominance relations among the group members. They seem to be aware that if individual A is stronger than B, and B is stronger than C, then A is stronger than C. In the wild, many primates behave according to such rules; they know who is dominant over whom, even when the groups in which they live are big (and for big groups it is quite unlikely that all dominance relations are stored separately in memory, because this would be a very long list). In the lab, many animal species were tested with artificial dominance relations: they would be trained with a set of stimuli, in which stimulus A yielded a bigger reward than B, and B a bigger reward than C. Many animal species (like macaques, rats and pinyon jays) showed that they were able to infer, after being trained to choose A over B and B over C, that they should choose A over C.[12]

This behaviour can be seen as the root of human transitive inference. However, Hurford notes, there is a lot of discussion, especially around lab experiments in this domain, and in general, 'transitive inference' in the case of animals is much more domain specific than it is in the case of humans (Hurford, 2007, p. 45–49).

After showing several respects in which animals approach human cognitive capacities, Hurford goes on to claim that animals (at least mammals) represent the world using what we could call *proto-propositions*. In the second half of the book, Hurford examines the evolution of communication. He notes that:

> Books necessarily have linear structure, and the separation into private and public evolutionary developments is an expository convenience, not implying that the evolution of private conceptions and of social behaviour were not intertwined; they certainly were.

Despite this remark, Hurford's sketch of evolutionary history is very much in line with Gärdenfors's, in the sense that Hurford lays an emphasis on the emergence of complex cognitive structure in animals. It seems that in his picture of the emergence of complex meaning, the private comes before the public. In other words, in primates, cognitive structure is there, but it is not used in communication.

> We have seen that animals lead quite rich cognitive lives, with signs of the beginnings of much of what has often been taken to be distinctively human. But mostly they keep this rich content to themselves. (Hurford, 2007, p. 164)

[12] In many studies the lists of items were longer than 3, ranging from four or five up to fifteen (Hurford, 2007, p. 45–49).

Hurford is not the only one implying this order of events in evolutionary history. A similar stance is taken by Fitch (2010). When the latter sketches the properties of the Last Common Ancestor of humans and chimps, he emphasises its cognitive capabilities and, like Hurford, states that its communication was limited.

> The LCA [Last Common Ancestor] had a rich suite of conceptual tools, forcing the conclusion that our shared ancestor could form concepts and memories, exhibit goal-directed behavior, draw basic inferences such as transitive inference, predict the movement of invisible objects, and learn complex serial orders, including at least some involving hierarchical tasks. Our prelinguistic hominid ancestors were sophisticated tool users. They lived complex social lives, requiring elaborate cognitive representations, and had well-developed concepts of space, time and causality. [...] In sharp contrast, [...] the communication systems of modern primates, and thus of the LCA, are far more limited. (Fitch, 2010, p. 173)

3.5.2 Criticising the cognitive road to meaning

Bickerton (2008) In a review of Hurford (2007), Bickerton (2008) characterizes two possible problems one can run into when talking about the emergence of language:

- The lower you rate the cognitive capacities of apes, the harder it becomes to explain how we (humans) got language.
- The higher you rate the cognitive capacities of apes, the harder it becomes to explain why they *did not* get language. (Bickerton, 2008, p. 285)

Hurford, Bickerton goes on to claim, is very much aware of the first problem, but thereby does not take the second seriously enough. To be more specific, according to Bickerton, Hurford gives animals the benefit of the doubt in too many cases, thereby positioning them too closely to humans in terms of cognitive abilities. When working from this position, it naturally becomes hard to address the second possible problem, but also the claim itself (that animals approach human cognition) has been a recent subject of criticism, for example in Penn *et al.* (2008).

Penn et al. (2008) In Penn *et al.* (2008), the claim is defended that there is a discontinuity between animal and human minds. Citing work from behavioral biology and psychology, the authors compare the cognitive capacities of humans and animals in different domains, such as causal and spatial reasoning and transitive inference. When they treat the latter topic, they give, like Hurford, an overview of research concerning serial ordering and transitivity, both in the lab and in the wild. They even report observations of fish in the wild being able, after seeing a series of pairwise fights, to keep track of a dominance ordering

among five rival males. Still, after presenting these and other results, they claim the following:

> [N]one of the available comparative evidence suggests that nonhuman animals are able to process transitive inferences in a systematic or logical fashion, even in the social domain. (Penn *et al.*, 2008, p. 116)

Their motivation for this claim is that the domains in which animals display this capacity are very much restricted:

> [T]he comparative evidence accumulated to date is nevertheless consistent with the hypothesis that nonhuman animals' understanding of transitive relations is punctate, egocentric, non-logical, and context specific. (Penn *et al.*, 2008, p. 117)

In an inquiry after an explanation for the discontinuity between animals and humans, they arrive, first of all, at the conclusion that *language* is not the factor that made the difference (Penn *et al.*, 2008, p. 120–123). To give a very short account of their arguments for this: they suggest that the kind of hierarchical structures one needs to speak complex modern languages most likely did not develop solely for language and that there are good reasons to assume that these hierarchical structures developed in cognition before language emerged. Thus, the crucial difference between animals and humans should be looked for in the domain of cognition, and eventually described in cognitive terms.

To explain the cognitive discontinuity between humans and animals, they propose the *relational reinterpretation hypothesis*. In short: "humans alone possess the [...] capability of reinterpreting [...] perceptually grounded representations in terms of higher-order, role-governed, inferentially systematic, explicitly structural relations" (p. 127). Penn et al. do not provide a fully worked out account of this hypothesis, so let me summarize at this point what is important in the context of this chapter: Penn et al. strongly argue against (a) the cognitive continuity between animals and humans and (b) the identification of language as the crucial difference between animal and human.

As was also noted in Bickerton (2008), the conclusions in Hurford (2007) and Penn *et al.* (2008) are based partly on, *the same* empirical research results. This shows that it is not straightforward to interpret empirical data and put them in a broader context, as is done in these publications.

If we focus on the specific example of 'transitive inference' in animals and humans, we see that Hurford is willing to describe what he sees (or what is reported by the researchers he quotes) in animal behavior as cases of actual transitive inference. Penn *et al.* (2008), on the other hand, are not. Even though they are looking at the same body of data, they draw different conclusions. Partly, I suspect, this difference is connected to whatever the authors *define* as 'being a case of transitive inference'. It seems that Hurford is more willing to ascribe the concept to what he sees in animals than Penn et al.. This means

that it is not only an empirical issue that is at stake here, but also a definitional one: is the term 'transitive inference' a term we should reserve for humans only, or should we loosen the term a bit and allow what is going on in some animals as cases of transitive inference too?

To recapitulate: the two publications I described—Hurford (2007) and Penn *et al.* (2008)—both provide analyses of and conclusions about empirical research on the behavior of animals. The disagreement between Penn et al. and Hurford illustrates that what is sometimes taken to be empirical evidence for one author, can be used as empirical evidence for the opposite view by another. This illustrates that for a uniform interpretation of empirical data, it is important to have well defined concepts, especially when the behaviour of both humans and animals is concerned. Moreover, although it seems intuitive to endorse the view that cognition, in evolutionary history, matured before language arrived, that is not necessarily the actual order of events.

3.5.3 Conclusion

We have seen an evolutionary picture sketched by Gärdenfors, in which the connection between cognitive structures and natural language meaning is deemed important, and in which, in the course of evolution, sophisticated cognition was in place before language became more complex. The evolutionary assumptions made by Gärdenfors can also be seen in the work of other authors focusing on the emergence of meaning, and I have called this common view *the cognitive road to natural language meaning*. We have also seen that this view faces problems: first of all, because the higher you rate the cognitive capacities of apes, the harder it becomes to explain why they did not get language. Secondly, because the empirical evidence that should prove that the cognitive capacities of various animals are sophisticated is not easy to interpret. In the literature, we have seen that behavioural studies that have been used to show complex traits in animals, have been used by different researchers to show that the behaviour under discussion is not that sophisticated, or at least very context specific.

I do not pretend to know the unique correct way to interpret studies of animal behaviour, nor do I have any idea how smart our closest cousins the chimpanzees, or any other species, really are. But I do think we need to be cautious, both when ascribing certain cognitive capacities to animals, *and* when saying that they do not have these capacities. The terms that are used to describe capacities are simply not clear enough to make such claims: what exactly is reasoning or transitive inference when it comes to species other than our own?

I am convinced that the cognitive road to meaning is a viable one, despite the problems described above. For many centuries, it has not been recognised at all that animals have capacities that approach human capacities in any way. The fact that animal cognition is now taken seriously as a discipline has opened

a lot of doors in the language evolution debate.[13]

On the other hand, we have observed that a characterisation of the transition from animal to human with too much emphasis on cognition might be unbalanced. To sum up, Gärdenfors's view that natural language meaning and cognition are closely connected is reflected in his sketch of the transition from animal to human, and this picture has been subject to criticism. In order to arrive at a balanced view let me recapitulate the first part of this chapter. We have observed that there are roughly three intuitions about the nature of meaning in natural language: first of all, the intuition that words and sentences refer to things in the world (referential theories of meaning). Secondly, the intuition that when a speaker makes an utterance, he typically intends his audience to come to believe something (meaning in terms of intention and belief). Thirdly, the intuition that our utterances have the meaning they have because these utterances have been used by other speakers in similar situations (conventional meaning).

So far, in this chapter, we have seen that the first intuition is not a good starting point for an evolutionary account of meaning (see section 3.2). The second intuition is the one underlying Gärdenfors's work, of which we have seen that it might need to be adapted in order to become more balanced. The third intuition is what we will turn our attention to next, because I suspect that when the conventional and communicative character of natural language is taken more seriously, this might result in a more balanced picture of its emergence. In other words, we will deviate from the cognitive road to meaning to see where a more communicative road leads us.

In the next section we will look at the work of Ruth Millikan, who, like Gärdenfors, introduces terms to talk about cognition and communication in animals and humans, and about the transition from animal to human. Unlike Gärdenfors, Millikan does not focus so much on cognition, when she discusses meaning, but rather relates meaning to *convention*. After describing Millikan's work, I will compare the two roads to natural language meaning and sketch how these two scenarios will be used in the remainder of this dissertation.

3.6 Meaning and communication

3.6.1 Millikan: Pushmi Pullyu Representations

In order to be able to describe animal behavior in a way that is useful in the context of the language evolution debate, Millikan introduced the notion of a *pushmi pullyu representation* (PPR). A PPR is a primitive kind of representation that conveys descriptive and directive information at the same time. In order to give a clear picture of what Millikan means by directive and descriptive information, let me quote a passage that she uses to illustrate it:

[13] See (Fitch, 2010, p.144–147) for a brief historical overview of this development.

> A list of groceries, Professor Anscombe once suggested, might be used as a shopping list, telling what to buy, or it might be used as an inventory list, telling what has been bought (Anscombe 1957). If used as a shopping list, the world is supposed to conform to the representation: if the list does not match what is in the grocery bag, it is what is in the bag that is at fault. But if used as an inventory list, the representation is supposed to conform to the world: if the list does not match what is in the bag, it is the list that is at fault. (Millikan, 2005, p. 166)

The first example, with the shopping list, is an example of directive information. The list represents, or directs, *what is to be done*. The second example, with the inventory list, is an example of descriptive information. Here, the list represents *what is the case*. As pointed out above, a PPR combines directive and descriptive information, but is more primitive than either directive or descriptive representations. To put this in Millikan's words, PPRs tell "in one breath both what is the case and what to do about it" (Millikan, 2004a, p. 18).

Although PPRs convey two kinds of information, they are highly inarticulate and are only fit for very specific tasks. They are only useful in situations where some kind of description and some kind of desired action is appropriate *at the same time*. Let me illustrate this with the following: an example of a PPR is what happens in the mental realm of a mother bird when it sees the open beaks of its young. When this happens, the mother gets descriptive information: 'these are the open beaks of my young'. But it also gets directive information: 'put food into these beaks'. The two kinds of information are present in the PPR at once; the mother cannot see the open beaks of the young without getting both the descriptive and the directive information. In other words, when the mother bird sees her young she cannot do anything but have the urge to feed them. Moreover, the two kinds of messages are in no way articulated in the representational state of bird. So, although PPRs are called *representations*, they are not representations in any classical sense of the word.

Millikan applies her notion of PPR not only to *mental representation*, but also to *communication*. An example of the latter is what is represented in the food call of a hen: when a hen sees food, she makes a particular sound to call her chicks over to the location of the food. Here we see as well that both directive and descriptive information is conveyed: in one undifferentiated breath, the hen says 'there is food here' and 'come over to get it'. Notice that when this kind of message is conveyed, the directive and the descriptive part play an equal role: it is not the case that one can be 'deduced' from the other. Moreover, as mentioned above, the directive and descriptive parts of the PPR are not articulated; when a PPR is communicated, both parts are transmitted *directly*.

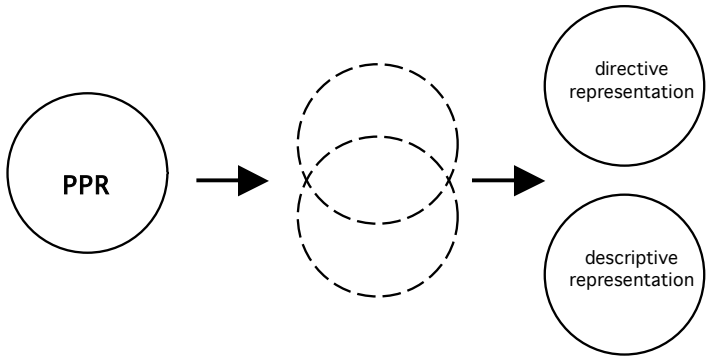


Figure 3.6: The emergence of human thinking, according to Millikan.

3.6.2 From PPRs to modern language

PPRs allow animals to react appropriately in numerous situations. The kind of behavior animals display may even look quite advanced, for example in the behavior of a cat who washes its ears by rubbing them with its paws, which it then licks clean. But at the same time the behavior is not flexible, and cannot be applied in different situations. So, the same cat does not know how to use its paws to "clean out the yummies at the very bottom of an emptied can of cream" (Millikan, 2004b, chapter 13).

More flexible behavior is only possible when an organism is able to understand a perceived situation without having to act on it. Then, the descriptive information that is used in one situation can be transferred to other situations. Millikan describes the evolution of more sophisticated organisms in exactly this way: she describes the transition from animal to human as a process in which PPRs get decomposed into their directive and descriptive parts (see figure 3.6). When an animal becomes able to take in descriptive information without necessarily carrying out a certain pattern of behavior, it starts to become able to represent things without directly using them. And thus it becomes able to mentally *simulate* different situations. To apply this to the cat example: if the cat were able to comprehend the technique it uses for washing its ears, without *actually* washing its ears, it could start to apply this technique in a situation where there is an almost empty can of cream. Thus, decomposing PPRs and storing various instances of descriptive information in order to apply them in different situations allows an organism to behave more and more flexibly.[14]

A crucial difference between humans and animals is thus, according to Mil-

[14]In Millikan (2004a), Millikan puts forward the hypothesis that the ability to decompose PPRs ran parallel with the evolutionary development of the forebrain and the division of what are called the 'dorsal' and the 'ventral' streams (p. 22).

likan, that humans have a lot of descriptive information that they cannot and need not apply directly. Although humans have developed these kinds of representations that go beyond PPRs, they still make use of PPRs, both in communication and mental representation. This situation is common in evolution: old mechanisms are typically not entirely replaced by newer ones. An example of a PPR in human mental representation would be the eye-blink reflex (see Millikan (2004b)): if something approaches the eye, the information *that something is approaching the eye* is represented at the same time and indivisibly together with the directive information *that you should close your eye*. An example of a PPR in human communication is a smile. Smiles tell us that "something potentially rewarding has just been done and to keep doing it" (Millikan, 2004b, chapter 13). But PPRs, Millikan claims, also occur in human language, such as in an English sentence like 'We don't eat peas with our fingers' (Millikan, 2005, p. 179): an expression like this describes and prescribes at the same time. In this case, the prescribed information is the same as the described information (namely, that we don't eat peas with our fingers). Note that in this example, like in the cases described above, the directive and descriptive parts of the message are transmitted *at the same time*. So the mechanism that plays a role here is, according to Millikan, not a Gricean implicature (Grice, 1975): it is not the case, according to Millikan, that the directive part is deduced from the descriptive part by the receiver, using conversational implicatures.

Given the concept of PPR, how do we describe the honeybee dance? Here the term can be put to use directly: what is represented in the honeybee dance is a PPR. When a honeybee dances to transmit information about a food source, it puts across two messages at once. First of all, that there is nectar at a certain location X, and, secondly, that the receiver of the message should go there to get it. The two parts of the message are not articulated, and cannot be transmitted separately; the receiving bee cannot observe the dance without actually obeying the order. This analysis of the honeybee dance differs in an important respect from Gärdenfors's: in his terminology, the honeybee dance refers to a cued representation of the food source. The fact that the observer of the dance 'should go over to this food source' is not part of the message, and should be 'inferred' by the observing honeybee (see also section 3.7.2).

Working from within Millikan's conceptual framework, what can we say about the chimpanzee, a species that is closely related to us in genetic terms but has much less sophisticated means of communication? We do not find a direct answer to that question in Millikan's work. Although she recognizes that in the higher animals, representations start to emerge that are less PPR-like, and start to look more like 'descriptive' representation, she also points out that much of the knowledge that is stored by these animals is connected to future use. In Millikan (2004a), she points out three things that humans are able to do, but that she suspects animals are not:

- To represent pure facts that concern situations or objects of a sort that have not yet proved to be of use either to the animal or to prior members

of its species.
- To represent facts about world affairs that have entirely unknown relations to the animal.
- To be motivated by representations that do not originate from the animal's perception of its current needs and/or current environment. (Millikan, 2004a, p. 26)

In other words, when thought becomes less PPR-like, representation is less connected to action. A representation can be stored just for the sake of representing, and it can be connected to an action later. Concerning the third item in the list, the claim that animals cannot be motivated by representations that do not originate from current perception, Millikan notes that

> Even our most respected and intensively studied relatives, the monkeys and apes, seem to derive their motivation entirely from perception of the current situation. (Millikan, 2004a)

In contrast, humans know all kinds of things that they don't have any particular use for, either now or in the future. And in order to learn these things, they do not need to be motivated by anything in the current environment. But that is not the only difference between animals and humans: we not only represent facts we don't know the use for; we also *communicate* these facts. And of course, there is a connection between the property of mentally storing so many facts and having a sophisticated public language. But the direction of this relation is not necessarily such that the fact that we store facts explains that we have language. On the contrary, the fact that we have language, according to Millikan, changes the way in which we perceive the world drastically:

> [A] very large portion of our [human] conceiving is done mainly or entirely through the medium of language (Millikan, 2005, p. 104).

I will not go into the discussion about what it means to perceive directly through the medium of language, but one thing is important here: according to Millikan, the fact that we have language makes a crucial difference for the way we perceive the world. Since chimpanzees do not have a public language in the way humans do, their mental capacities are of a different order than those of humans. Thus, Millikan has much less confidence than Gärdenfors in the mental capacities of chimps. This difference between Gärdenfors and Millikan, and some others we hinted at in this section, will be the topic of section 3.7. But let us first look at what Millikan says about meaning in human language.

3.6.3 Millikan on meaning in modern language

Contrary to both Gärdenfors and Grice, Millikan does not see a direct connection between the content of an utterance and what is mentally represented by the speaker of the utterance:

> Linguistic meaning of a public language makes no *direct* reference to thought. (Millikan, 2005, p. 96)

Instead, according to Millikan, it is important to distinguish between two kinds of meaning: the meaning a speaker has in mind when he communicates and *conventional meaning*:

> Language is a tool used in communication; linguistic meaning is derived, somehow, from the ways in which people use linguistic devices in intentional communicative acts. Still, in finding ways to communicate, we do not need to start from scratch each time. There develop certain linguistic conventions—recognisable patterns in the way in which people use linguistic expressions and grammatical constructions in order to communicate. And so a distinction emerges between the meaning that a speaker intends to convey on a particular occasion, and the meaning conventionally associated with the words that the speaker has used. (Price, 2007, p. 766)

Here it shows that Millikan's view on meaning is influenced by Lewis's account of conventional meaning (the one we have seen a brief description of in section 3.2.4). Conventions, Millikan claims, are generally patterns of behaviour that are reproduced, and (some of these) conventions regulate coordination. Coordination is achieved when two individuals' combined actions achieve a goal together, and conventions help to streamline behavioural patterns so that coordination is achieved more easily. This description of convention applies to language (we choose certain utterances in certain situations because they serve the communicative goal we had in mind), but also to very basic behavioural patterns like a handshake (when we shake hands upon meeting someone, we know exactly how to move our hand in order to make the handshake work properly; this involves harmonising the arm movements of both 'handshakers') or moving something heavy, like a sofa, together (in such situations, usually one of the two movers takes the lead, and the other adjusts his movements in order for the moving to run smoothly). This characterisation of coordination shows that Millikan pictures conventions as something very much entrenched in behaviour, something that can work unconsciously. The knowledge needed to control such coordination, knowledge of conventions, can be stored in a very primitive form like PPRs (Millikan, 2005, chapter 5). Thus, Millikan shows that conventions are not a product of purely rational minds (and this sets her position apart from Lewis's; see (Millikan, 2005, chapter 1)), but instead they are deeply rooted in behaviour in general, and it is easy to see that even animals coordinate their behaviour in this way on some occasions. As an example of this, Millikan mentions intention movements in the behaviour of animals as described by Tinbergen (1952). Intention movements are movements from which another animal can infer what is likely to be happening next. These movements can be useful to animals, for example, when by making an intention movement for fighting, they can actually prevent real fighting from taking place.

Public languages are just many conventions thrown together:

> A public language is a huge web of crisscrossing lineages of reproduced patterns consisting of tokens of linguistic forms and responses to them. [...] Words, idioms, syntactic forms, tonal inflections, and so forth are handed down from one person to others because these elements are helping to serve coordinating functions. (Millikan, 2005, p. 60)

It is not necessary for a convention to be followed at all times by all individuals. I can secretly take a certain syntactic expression to mean something I made up myself, or understand someone else's utterance in my own way. What I have in mind when I say something is less important, according to Millikan, than what my utterance means according to the running convention.

With this statement, we are where we started this section: I wanted to illustrate that, according to Millikan, linguistic meaning is not directly connected to a speaker's intention, but is rather something conventional. Conventional meaning can be understood in a Lewisian manner, but Millikan's account differs from it in that she sees conventions much less as something purely rational, and much more as something rooted in behaviour in general.[15]

It is clear that both the evolutionary scenario Millikan sketches and the account of meaning she has worked out are different from Gärdenfors's, but their work also shows similarities. In the next section we will focus on these similarities and differences.

3.7 Two roads to language

The preceding sections introduced us to the work of Gärdenfors and Millikan. Both have worked on describing the emergence of language, with an emphasis on meaning. To make this description, they introduced new concepts, that aim to describe a gradual transition from animal to human, and they sketched evolutionary scenarios of this transition. Let me recapitulate.

Gärdenfors introduced the concepts of 'cued' and 'detached' representation. He described the transition from animal to human as a process in which organisms are first able to form detached representations, subsequently form an 'inner environment' and then start to communicate using sophisticated structure. We called this view *the cognitive road to language*, because it postulates that sophisticated cognition was there before sophisticated communication, and it is compatible with a view of natural language meaning that explains meaning in terms of mental representations.

[15] See Rescorla (2011).

Millikan introduced the concept of 'pushmi pullyu representations' (PPR) and described these as primitive representations that include information both about what is at stake and about what must be done. PPRs play a role in both cognition and communication. The trajectory from animal to human involves decomposing PPRs into separate descriptive and directive representations, which makes it possible to divorce descriptive information from action. Natural language meaning is pictured by Millikan as something essentially conventional, and she shows that these conventions are deeply rooted in behaviour in general, thus providing a biological account of convention.

From this description it is clear that there are many differences between Gärdenfors and Millikan, but let us first focus on something they share.

An important common emphasis in the accounts of Millikan and Gärdenfors is the emphasis on the detachment of representations from the here and now and the directly necessary. Both authors stress the importance of representing things without actually carrying out an action that is connected to the things represented. This makes organisms able to test hypothetical situations, and thereby their behavior becomes more flexible: they can store knowledge for future use, without necessarily having to act.

This common emphasis is interesting because it shows that both authors recognise that humans are more ready than other species to store knowledge without using it immediately. They do take different positions, however, when it comes to an exact characterisation of the differences between humans and other animals, as well as the order of events in the evolutionary history of human language and the properties of meaning. Let us have a look at these differences.

3.7.1 Different evolutionary scenarios

Whereas Gärdenfors defends the hypothesis that chimps have a cognitively rich life but do not have language, Millikan claims that chimpanzees, like other animals, are mainly motivated by factors that are present in the immediate environment and that they differ crucially from humans in this respect.

If we look at these claims in an evolutionary setting we can conclude the following. Millikan's line of thinking is compatible with a scenario in which the development of certain cognitive capacities that humans have now emerged only after the development of a sophisticated communication system, whereas Gärdenfors's ideas are more in line with a scenario where a great deal of the cognitive capacities have matured before language emerged.

When we focus on more general questions about the relation between language and thought, and the role of this relation in evolution, we see differences between Gärdenfors's and Millikan's approach as well. Millikan stresses that human mental capacities are at least in part dependent on linguistic capacities.

Gärdenfors, on the other hand, emphasizes that complex mental capacities are possible without linguistic capacities.

Of course, the two positions are not mutually exclusive. Gärdenfors does not deny the influence of language on modern human thought and Millikan does not claim that it was the emergence of full modern language *alone* that made human thinking more sophisticated. What I have sketched here is a difference in emphasis.

3.7.2 Different views on meaning

Both Millikan and Gärdenfors have not only written about meaning in an evolutionary setting, but also about meaning *per se*. In other words, both have made claims about the nature of natural language meaning, and we have seen short descriptions of these claims in the preceding sections.

The main difference between the two at this point is that Gärdenfors claims that the content of an utterance corresponds directly to a mental representation. According to Millikan the relation between the two is mediated by a linguistic convention.

When it comes to the connection between an utterance in language and a mental representation present in the speaker, Millikan postulates a much looser connection than than Gärdenfors. This becomes clear when we focus on what they define as communicative success. Gärdenfors would claim that there is communicative success when the hearer recognizes the belief that the speaker intends to express, whereas Millikan defines success as the cases where the hearer recognizes the linguistic conventions the speaker had been following.

Let us have a look at the honeybee dance again. According to Millikan the dance transmits two messages in one breath; a descriptive and a directive act takes place at the same time. According to Gärdenfors, the dance is a signal that refers to a cued representation of the food source. And because a cued representation is a representation of something that is close-by or triggered by something that is close-by, the dance can be seen as representing the food source itself. That the receiving bee should go over to the food source to get the food, is something that is not conveyed in the message. It is a conclusion that is drawn by the receiving honeybee itself.

The latter kind of analysis seems quite parallel to what Grice (1975) describes as going on in language use: receivers make—consciously or unconsciously—inferences from the received message and thereby decide what the full meaning of the message is.

Thus, we noted that the way Gärdenfors would characterize honeybee communication fits well in a Gricean school. Millikan's analysis of the honeybee dance, on the other hand, is much less compatible with a Gricean approach. According to her, no 'inference' takes place in the receiving honeybee; the directive part of the message is properly part of the message, and is interpreted in a very direct way.

3.7.3 A cognitive and a communicative road to language

Above, we have connected Gärdenfors's view on the emergence of language to a general attitude towards the emergence of language, described as *the cognitive road to language*. It was pointed out in section 3.5.1 that many authors seem to have the same conviction as Gärdenfors, that complex cognition developed (to some extent) prior to complex language, and that linguistic meaning should be explained in terms of cognitive structures.

We have also seen, in section 3.5.2, that this picture faces problems, and that it would be attractive to consider other views, in order to lessen the emphasis on cognition and create a more balanced picture of the emergence of language and meaning.

Let us see if the picture sketched by Millikan could fulfil this function. As described above, Millikan puts less emphasis on cognition than Gärdenfors when it comes to describing meaning in natural language. Instead she stresses the conventional nature of meaning. When describing the evolutionary trajectory of the emergence of language in humans, she uses the same concept (PPRs) for describing both *communication* and *cognition*. Moreover, she stresses that having language has changed the way we see the world enormously.

Because of the role she ascribes to communication, and the role that conventions (which are much more principles of communication than of cognition alone) play in her description of linguistic meaning, I will call her view of the emergence of language and meaning in language *the communicative road*.

Now we have two 'roads to language', it would be natural to claim that one of the two is a better view. I will not do that, however, because I think that both views capture aspects of language and language emergence that are important, and choosing sides would only lead us away from a balanced picture of the emergence of language.

Instead of choosing sides, I formulate the following take home message: in the literature, there is much emphasis on a view like the cognitive road to language, but in order to arrive at a more balanced picture of the emergence of language, and to account for the conventional nature of language, we need to take into account the communicative road to language as well.

Thus, I am not going to choose sides, but neither am I going to devise a philosophical account that should serve as a combination of the two pictures. I think we have seen enough concepts and abstract proposals by now, and I would like to focus, again, on empirical data. In the next chapter, empirical data that should tell us something about protolanguage is discussed, and I will discuss how the two roads to language presented here can help us to interpret the data in a justifiable manner.

The accounts of meaning discussed so far in this chapter mainly discuss the properties of meaning *in general*; they look at meaning as a general phenomenon. But if we want to arrive at a complete overview of what has been said about meaning, we should also look at the meanings of our utterances in a more detailed way, by looking at what happens when words are put together into

sentences and complex meaningful structures are formed. This is how linguistic semanticists generally approach meaning, and we will follow this approach by looking at two phenomena in which structured meaning plays a role: compositionality and thematic relations. This will equip us with the terminology necessary to analyse the utterances made in restricted linguistic systems.

3.8 Meaning in linguistics: the structure of meaning

As stated above, linguists take meanings of utterances to be the point of focus, rather than meaning as a general phenomenon. They take as a starting point the assumption that sentences or utterances express propositions, and they study the structure of these propositions, possibly in a formal modeling context.

In this section, I will discuss two concepts that can help us analyse meaningful structures. First is the concept of **compositionality**, a term that is central to the discipline of formal semantics. Semanticists construct mathematical models of natural language, in which they assume (at least, this is done in many existing formal models) that the meanings of words, together with the way they are put together, make up the meanings of sentences. Secondly, I will discuss the notion of **thematic relation**, a term that one can come across in syntactic theory and in semantics (but not so much in model theoretic semantics), and that categorises arguments of verbs into different classes, each class having its own special characteristics.

As I have said, both concepts ('compositionality' and 'thematic relation') have something to do with the formation of complex meaning in natural language, and they will be of importance in the last three chapters of this dissertation.

3.8.1 Compositionality

Formal semanticists construct models of natural language, in order to describe the meanings of utterances with logical and mathematical means. Formal semanticists might start from a general account of natural language meaning, but for their models, they look at language in a more detailed way: they investigate how utterances get their meaning. Many semanticists take the principle of compositionality to be a leading notion. This notion is generally defined as follows:

> The meaning of a complex expression is determined by the meaning of its constituents and the way they are put together.[16]

[16] See Szabó (2008) for a similar definition, and an overview and discussion of related definitions.

It was observed in Chapter 1 that looking at the origins of meaning in human language is most interesting if we do not take meaning to be only a property of words, but something that has complexity: complex meanings can be built using language. The principle of compositionality captures this intuition.

Formal semanticists generally take care that their models of natural language are compositional as well: these models build the meanings of sentences by taking the meanings of individual words and combining them, using information about the syntactic structure of the sentence.

As an argument in favour of the compositional nature of language, it is often pointed out that language is *productive*:

> **Argument from productivity:** Since competent speakers can understand a complex expression e they [have] never encountered before, it must be that they (perhaps tacitly) know something on the basis of which they can figure out, without any additional information, what e means. If this is so, something they already know must determine what e means. And this knowledge cannot plausibly be anything but knowledge of the structure of e and knowledge of the meanings of the simple constituents of e. (Szabó, 2008)

Another argument in favour of compositionality is the argument of systematicity. This argument says that anyone who understands the expression 'the rug is under the chair' also understands 'the chair is under the rug'. In other words, there is systematicity in the way we understand sentences. The only way in which this could be the case is that the meanings of complex expressions depend systematically on their constituents and their structure.

These two arguments show that natural language must be compositional, at least to some extent, but there are cases where natural language does not seem to be entirely compositional. For example, do all who understand 'within an hour' and 'without a watch', also understand 'without an hour' and 'within a watch'?[17]

Most formal semanticists recognise that language is to some extent compositional, but that not all of natural language is compositional. Some even state that there are no *empirical* reasons to model natural language compositionally, but that we can choose to model meaning compositionally, because it leads to elegant theories (Groenendijk and Stokhof, 2005).

Thus, it can be seen that semanticists do not fully agree on whether compositionality should be seen as an empirical property of language or as a starting point taken for modeling reasons. The take home message that I would like to distill from this discussion about compositionality is that language clearly shows the properties of productivity and systematicity at least to some extent. Thus, when we focus on the meanings of sentences, we need to take into account that they are at least partly determined by their constituents and the way these

[17] The example is from Szabó (2008); it is mainly an argument against the systematicity of language. The same publication also lists arguments against the principle of compositionality.

constituents are put together. An interesting evolutionary question is how this systematicity came about. In short: how did the property of compositionality come about?

3.8.2 The emergence of compositionality

Two opposing views of the emergence of compositionality can be found when we analyse discussions about protolanguage. In Chapter 1, several accounts of protolanguage were described, and it was noted that there is an opposition between holistic and synthetic protolanguage. Those proposing that protolanguage was holistic hypothesise that utterances with unanalysed structure were uttered and, through usage, chopped up into a systematic system. Those arguing in favour of a synthetic protolanguage (such as e.g. Bickerton) hypothesise that protolanguage consisted of units that were very much like words in full language and that were put together into sentences. These two views indirectly say something about compositionality. Proponents of holistic protolanguage claim that language became compositional by *decomposition*: unanalysed multi-syllable words were decomposed into analysed, complex structures with compositional meaning (this process is described in, among others, Kirby (2000); Wray (1998)). As described in Chapter 1, Wray illustrates her account of protolanguage with two imaginary utterances, /mɛbita/ and /kamɛti/, which are associated with the meanings *give her the food* and *give me the food* respectively. After analysing the analogy between the shared syllable /mɛ/ and the shared meaning 'singular female recipient', speakers of holistic protolanguage will take this regularity into their inventory (Wray, 1998, p. 55). With segmentation happening bit by bit, phonological units without an ascribed meaning will be omitted and a compositional language is the result. The process that Wray described, segmentation of unanalysed strings into a compositional language, was replicated with computational models (Smith *et al.*, 2003; Kirby, 2000), as well as with participants in an iterated learning experiment in the lab (Kirby *et al.*, 2008).

To connect the concepts of compositionality and protolanguage to Gärdenfors's and Millikan's views presented in the previous section, the basic idea behind holistic protolanguage is found in Ruth Millikan's work as well:

> It is a serious mistake to suppose that the architectural or compositional meaning of a complex sign is derived by combining the prior independent meanings of its parts. Rather, the meanings of the various significant parts or aspects of signs are abstracted from the prior meanings of complete signs occurring within complete sign systems. (Millikan, 2004b, chapter 4)

Gärdenfors, on the other hand, portrays protolanguage as a synthetic system (Gärdenfors, 1995, section 5), a view of protolanguage which I will discuss next.

Proponents of synthetic protolanguage claim that protolanguage already had a compositional character (but this was not necessarily fully developed

yet). In recent literature, synthetic protolanguage is called 'compositional protolanguage'. We will focus on two such accounts to see how the emergence of compositionality is depicted.

In both Jackendoff (2002) and Hurford (2011), a two step procedure is depicted that led to compositional structure in language: first concatenation of different meaningful units into proto-utterances, then more systematic usage of word order. Let us have a closer look at how they sketch this process. Hurford distinguishes a stage in the emergence of language where utterances were *purely pragmatic*: words were put together, and the relation between these words was to be retrieved from the context. In this stage, he claims, there is already a rudimentary form of compositionality present:

> In the purely pragmatic mode, with no grammatical rules or conventions for putting words together (beyond mere concatenation), there is already a simple systematic relationship between the meaning of a whole string and the meanings of its constituent words. For example, the meaning of *Mommy sock* is something to do with Mommy **and** something to do with a sock. The meaning of the whole is any proposition that a hearer can extrapolate from shared background knowledge and contextual clues, provided that this proposition is somehow about Mommy and a sock. The meaning of Daddy car is not the same as the meaning of Mommy sock, exactly because of differences among the meanings of Mommy, Daddy, sock and car.

Thus, compositional language was preceded by a stage in which words were put together, of which the meanings had something to do with each other. According to Hurford the meaning of such proto-utterances was to be derived by the hearer in this stage, because the order of the different words in the utterances did not contribute information (hence, this stage had only a 'rudimentary form of compositionality' and not full compositionality). Thus, what appeared first was the combination of meanings in a rough manner, without paying attention to word order. Jackendoff (2002) sketches a similar stage and calls this the stage in which 'concatenation of symbols' takes place.

The principles 'AgentFirst' and 'FocusLast' (described in Chapter 1.4.3) of Jackendoff can be seen as an answer to the question how we got from a rudimentary compositional stage into full compositional language: semantic and pragmatic principles determined the position of different words in utterances. Or, as Jackendoff puts it, the placement of symbols conveyed information about basic semantic relations. I will focus in more detail on these 'basic semantic relations' in the next section.

To conclude, when the properties of protolanguage are discussed, compositionality turns out to be a central concept, and two intuitions exist about the emergence of compositionality. According to those proposing a holistic protolanguage stage, compositionality emerged 'by decomposition'. According to those proposing a synthetic (or compositional) account of protolanguage, compositionality emerged 'by composition', and in two stages. In the first stage,

words were combined into multi word utterances, but the order of the words was not systematic. In the second stage, the position of the words was used in order to convey additional information. Thus, in the second stage, the meaning of an utterance is the combination of the meanings of the words *and* the way they are put together.

In this dissertation, we will focus in further detail on synthetic protolanguage, and particularly on the second stage mentioned above: how did word order in protolanguage utterances contribute to their meaning? In the next section, I will describe a phenomenon that will help us to describe this process: *thematic relations*.

3.8.3 Thematic relations

Virtually all languages have verbs and nouns. Verbs typically describe actions, and nouns persons or objects. Nouns can act as arguments of verbs: they denote participants in the action described by the verb. Actions can have different participants, and the possible ways to participate in an action are described as thematic relations.

Thematic relations describe the roles a participant in a certain action plays with respect to the action. For example, in a situation where *John throws a ball*, the action is throwing, and John and the ball are the participants in the action. John does the throwing, so his role is the role of *agent*. The ball is affected by the throwing, so the role of the ball is that of *patient*. Semantic role theories aim to provide a list of roles that can apply to any argument of any verb. When that is done, verbs can be analysed in terms of the arguments they take.

Inventories of thematic roles were proposed by researchers working on syntax and semantics and empirically studied by psycholinguists. In the literature, lists of thematic roles have been proposed that have, among others, Agent, Patient, Experiencer and Theme as well known members. The following is a list with descriptions of these thematic roles from Dowty (1989).

- *Agent* — a participant which the meaning of the verb specifies as doing or causing something, possibly intentionally. Examples: subjects of *kill, eat, hit, smash, kick, watch*.
- *Patient* — a participant which the verb characterizes as having something happen to it, and as being affected by what happens to it. Examples: objects of *kill, eat, smash,* but not those of *watch, hear and love*.
- *Experiencer* — a participant who is characterised as aware of something. Example: subject of *love, object of annoy*.
- *Theme* — a participant which is characterized as changing its position or condition, or as being in a state or position. Example: object of *give, hand*, subject of *walk, die*. (Dowty, 1989, p. 69)

These examples can be seen as representative examples of how thematic relations are formulated. However, there is no systematic account of thematic

roles, and there exists little agreement about which thematic relations exist and how they should be defined exactly,[18] even though some researchers do not recognise this and take the notions for granted.

An analysis of thematic roles that is still quite influential was put forward by Dowty (1991). Dowty proposes that different roles should not be seen as discreet categories, because making such a discreet categorisation will always run into problems. Instead he proposes that there are two proto-roles: Proto-Agent and Proto-Patient. Each proto-role has contributing properties. For instance, a Proto-Agent is typically volitionally involved in the event or state, whereas a Proto-Patient typically undergoes a change of state. A Proto-Agent typically causes an event or state in another participant, while a Proto-Patient is stationary relative to the movement of another participant.[19] These contributing properties do not necessarily all play a role in each predicate: in 'John is ignoring Mary,' John is volitionally involved, but he does not cause an event or state in Mary. Rather, the properties of the proto-roles can help us in determining which roles are associated with which grammatical relations. This proceeds in the following way:

> In predicates with grammatical subject and object, the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalized as the subject of the predicate; the argument having the greatest number of Proto-Patient entailments will be lexicalized as the direct object. (Dowty, 1991, p. 576)

With the help of the two proto-roles, Dowty successfully analyses a number of problematic issues in lexical semantics (Dowty, 1991). His work indicates that among the vast number of possible thematic relations, *agent* and *patient* are the most important. Let us have a look at what role these relations play in the language evolution debate.

3.8.4 Thematic roles in language evolution

To see thematic relations at work in an account of the emergence of language, we will focus on Jackendoff's work on protolanguage.

As we have seen above, Jackendoff sketches a picture of the emergence of language in which there are two important stages: first, words are put together, but no importance is given to the order in which they are put together. Subsequently, the position of the words in the utterances is used to signal additional semantic information. One of the rules that expresses these relations is Agent-First. This is a rule that tells us that the individual with the thematic role of Agent is put first in an utterance (see also the description in section 1.4.3).

[18] 'There is perhaps no concept in modern syntactic and semantic theory which is so often involved in so wide a range of contexts, but on which there is so little agreement as to its nature and definition as [...] thematic relation' (Dowty, 1991, p. 547).

[19] See (Dowty, 1991, p. 572) for the full list of properties.

Thus, it is suggested that in the protolanguage stage, information about thematic roles is used in the organisation of utterances. In other words, in a semantic protolanguage, thematic role is one of the sources of information on the basis of which utterances are organised (besides other sources of information such as e.g. 'newness of information,' as in the rule FocusLast). In the chapters that follow, several sources of information for utterance structure will be reviewed, and a more detailed picture of the organisation of utterances in protolanguage will be sketched, and empirical evidence for this more detailed sketch will be provided.

Above we have seen that the two proto-relations Proto-Agent and Proto-Patient are associated with the syntactic categories subject and direct object. Of course there are exceptions (some are discussed in Dowty (1991)), but in full language, it is almost always the case that an argument one would describe as agent in semantic terms is indeed also the subject of the sentence, and the patient is the direct object. Given Jackendoff's AgentFirst principle, it would be expected that many languages have a word order in which the subject comes first. This is indeed the case: a fair share of the languages of the world has either SOV or SVO as its dominant order. This phenomenon will be discussed in more detail in Chapters 4 and 5.

To conclude, thematic roles are one of the sources of information on the basis of which language systems without full syntax are organised. Their effect is still visible in restricted linguistic systems, and they can be hypothesised to have played a role in the emergence of language, as precursors of syntactic rules.

3.9 Conclusion

Throughout this chapter we have seen that natural language meaning is studied in many ways. Linguists and philosophers have engaged in discussions about the nature of meaning and about the way meaningful structures are built when we interpret utterances. It is difficult to distill one analysis from the overview that was given in this chapter, but we have come across concepts and ideas that are important in the analysis of meaning, and that might be important when we investigate the evolutionary history of meaning. I will now give a list of these concepts and ideas. We will use them in later chapters as 'tools for thinking' about semantic protolanguage. For clarity, I will display the most important concepts in boxes.

3.9.1 Defining meaning: reference, intention/belief and convention

This chapter started with a very brief overview of the theories of meaning that were put forward in the 19th and 20th century. Many theories of meaning do not account for the fact that language evolved; they simply look at language as

an existing system. Still, these different accounts told us something about what is important about meaning; we have seen three intuitions about meaning that should be taken into account:

reference The things words and sentences refer to are important when we want to know their meaning.

intention and belief When a speaker makes an utterance, he typically intends his audience to come to believe something.

convention Our utterances have the meanings they have because these utterances have been used by other speakers in similar situations.

The first intuition is the basis of propositional accounts of meaning, and these capture a very prominent fact about language: that we use it to refer to things in the world. This capacity is very prominent in humans, and it allows us to talk about things that are not present *here and now*. This property of language will be discussed further in chapter 6.

Further, a central problem for referential theories of meaning, that of intensional contexts, will show up later in this dissertation: in Chapter 5, I will show that intensional expressions possibly had a special status in protolanguage.

The second and third intuitions form the basis of two evolutionary scenarios of the emergence of language and meaning: the cognitive road and the communicative road to language. Thus, the overview of different accounts of meaning has not given us a theory to work with immediately, but has pointed out prominent facts about meaning in natural language.

3.9.2 Roads to language: cognition and communication

In the second part of this chapter, we focused on concepts that were introduced by philosophers in order to get a more satisfactory description of the transition in evolutionary history from animals to humans. These concepts focused on cognitive structures and on meaning in language or communication in animals and humans.

We observed that the concepts introduced by Peter Gärdenfors fit very well in a set of hypotheses about the evolutionary history of meaning in which cognition develops before meaning. This position is taken as a starting point by many other authors who have written about the emergence of meaning in human language.

We discussed a second framework of concepts, introduced by Ruth Millikan, which focuses less on cognition and emphasises the conventional properties of language, and compares language to other cultural phenomena: behavioural

patterns which are copied by individuals, like the habit of shaking hands when meeting in many western societies.

There is much to say for a Gärdenfors-like approach to the emergence of meaning, which emphasises the relation between cognitive structures in language users and their utterances, because it points our attention to the cognitive properties of our ancestors (animals). Cognitive capacities found in them might be ancestors of features of our linguistic capacities. Much interesting research has been carried out already to describe these links (Fitch, 2010).

The line of thinking introduced by Ruth Millikan, however, points our attention away from cognitive structures to conventions, to language as a public phenomenon. Millikan denies that there is a one to one relationship between the meaning of an utterance and what is mentally represented by the speaker or the hearer. By emphasising the conventional nature of linguistic meaning and grounding conventional behaviour biologically, she takes the communicative nature of language into account.

A cognitive road to meaning: an evolutionary scenario in which cognitive capacities are central, and are held responsible for the structure of the emerging language.

A communicative road to meaning: an evolutionary scenario in which the dynamics of communication and the formation of conventions are central.

In the next chapter we will continue to discuss the claim that the structure of utterances in protolanguage was governed by semantic and pragmatic principles. The two roads to meaning described here will help us to analyse the nature of this scenario: are the principles governing protolanguage essentially *cognitive* or *communicative*?

3.9.3 Building a proposition: compositionality, thematic roles and topic/focus

In the last part of this chapter I focused on the formation of complex meaningful structures in language. I described two phenomena, compositionality and thematic relations, and I described the role of both phenomena in the evolution of language.

First of all, I pointed out that two different claims have been made about the emergence of compositionality in natural language, and these connect to the accounts of protolanguage discussed in Chapter 1. More precisely, holistic and synthetic protolanguage each make different claims about the emergence of compositionality.

Holistic protolanguage: compositionality emerged by decomposition and reanalysis

Synthetic (compositional) protolanguage: compositionality emerged in stages, by composition

Secondly, I described the role of thematic roles in language evolution. There is a natural connection between the role of Agent and the syntactic subject, and the role of Patient and the syntactic direct object. Thematic roles are one of the sources of information on the basis of which language systems without full syntax are organised. Thematic roles can be seen as the precursors of syntactic categories. This will be discussed in more detail in the next chapter.

Agent → Subject
Patient → Direct object

To conclude, this chapter has equipped us with a set of useful terms and concepts to describe meaning and its origins. These terms and concepts will be put to use in the next chapter, where we will focus on restricted linguistic systems 'in the wild' and in the laboratory.

CHAPTER 4

Cognition and communication 'in the wild' and in the lab[1]

4.1 Introduction

In the previous chapter, we have seen a detailed analysis of meaning in natural language. In this chapter I will show how we can operationalise the concepts described there. To be more precise, I will show how we can use these concepts to collect stronger and more detailed evidence for a semantic protolanguage account. Let me summarise what I have said about semantic protolanguage earlier in this dissertation.

Chapter 1 proposed and developed a description of the semantic account of protolanguage: the hypothesis that semantics (and pragmatics) had a role in the structure of protolanguage. This hypothesis goes back to ideas advanced by Ray Jackendoff, and we have reviewed some of the sources of evidence he used to support his claim. Moreover, a brief overview of observations from different restricted linguistic systems shows that semantic properties could play a role in the organisation of these systems.

We concluded that this approach (looking at restricted linguistic systems) is promising, but that it would be good if we would have more data. We also observed, however, that evidence is not easy to collect, because restricted linguistic systems typically emerge in unfortunate and exceptional situations. Moreover, the step from restricted linguistic systems to evolutionary conclusions is still

[1]Part of the research presented in section 3 of this chapter was originally published as Schouwstra (2010).

unsatisfactory, because it is not clear how the analogy between restricted linguistic systems and protolanguage works.

In other words, in order to make the semantic protolanguage account stronger, we need to do three things.

1. Collect more data from restricted linguistic systems.
2. Develop a more detailed account of how we can draw evolutionary conclusions from data of simple language systems.
3. Describe in more detail the principles that may have played a role in semantic protolanguage.

In order to address issue (1), I will present a new way of collecting restricted linguistic data: by 'creating' restricted systems in the laboratory (in section 4.2). This method will allow us to collect data in a more controlled way than the data that is at hand now. Subsequently, in section 4.3, I will address issue (2) and specify how the data obtained in the lab can lead to evolutionary conclusions. In other words, I will sketch how the analogy works between the observed data and the evolutionary conclusions we can draw from them. I will do this by first discussing different existing bridge theories and formulating a bridge theory that serves my purposes. In this discussion of bridge theories, the concepts that were discussed in the chapter on meaning will play an important role.

From the discussion of meaning in an evolutionary context I distilled two possible views on the trajectory towards full language: the cognitive road and the communicative road. The former view stresses the fact that language sprouts from cognitive abilities of individual language users, and seeks an analogy between meaning in language and mental representations in speakers and hearers. The latter focuses more on language as a public phenomenon: it stresses the conventional nature of language. If language took the cognitive road, it is hypothesised, much of language matured in the cognitive domain before it was used as a communicative system. If language took the communicative road, its structure does not necessarily mirror cognitive structures, but communicative pressures are taken to be important. According to this view, communicative interaction is responsible for the structure of language. The distinction between the cognitive and the communicative road to full language is not meant as a rigid distinction; the two trajectories represent two complementary ways of looking at language. In fact, I will put forward the hypothesis that both cognitive and communicative principles played a role in protolanguage, and specify how they interacted (in section 4.3.4). This hypothesis will be tested in the next chapter.

In the last part of this chapter (in sections 4.4 and 4.5), I will address issue (3) from the list above and formulate new empirical questions that can be investigated: one about the emergence of word order, and the other about *temporal displacement*. I will look at data from existing restricted linguistic

systems to see if answers can be found there. After this is done, the ground is prepared for the two chapters following this one, where we will study the newly formulated questions (about the emergence of word order and temporal displacement) in an improvised communication setting.

4.2 Creating restricted linguistic systems in the lab

In chapter 2 we have seen that restricted linguistic systems (Basic Variety, pidgin, homesign and newly emerged sign languages) all arise in situations where 'normal' communication is not possible. Even though the circumstances for the individual systems are different (for instance, the situation of ABSL signers is quite different from L2 learners), the systems show striking similarities, and the data that are gathered may offer us a valuable peek into the workings of protolanguage.

We have also seen that the collection of data from restricted linguistic systems is not easy because these systems emerge in exceptional situations, and usually occur in unfortunate circumstances. Moreover, interviewing people in these situations does not give researchers optimal control over the kinds of constructions and utterances made by the speakers. Finally, in existing restricted linguistic systems, it is usually only language *production* that is studied, and if we want to take the communicative character of language seriously, it is necessary to look at *interpretation* of utterances as well.

In this section, I will describe a kind of laboratory experiment that can serve as a way to collect restricted linguistic evidence, and does not face the problems described above: the improvised communication task. Subsequently, I will reflect on evolutionary conclusions drawn on the basis of the lab data: what bridge theory do we need in order to establish firm ground for evolutionary assumptions made on the basis of the experimental data?

The improvised communication task was introduced by Goldin-Meadow *et al.* (2008).[2] They asked participants from four different language groups to communicate about simple events that were presented in short animations. The participants were not allowed to speak, and were asked to communicate using gesture. The participants were not familiar with any conventional sign language and they were encouraged to use improvised, iconic gestures to make clear what was depicted on the vignettes.

The events used as stimuli in the experiment were simple events in which some motion was involved ('motion events'). All vignettes depicted an Act (some action involving motion), an Actor (the performer of the Act) and a Patient (the undergoer of the Act).[3] Examples are 'man carries baby' or 'boy

[2]The term 'improvised communication' is not theirs, however; it was introduced by Langus and Nespor (2010). See also the descriptions of these two publications in section 5.

[3]Goldin-Meadow *et al.* (2008) also used vignettes with only an Actor and an Act, but we

tilts glass to mouth'. Note that Goldin-Meadow *et al.* (2008) use terminology similar to that of the thematic roles discussed in chapter 3 (Actor and Patient correspond to Dowty's proto-agent and -patient).[4] Recall from section 3.8.4 that there might be an evolutionary connection between the semantic classes Agent (Actor) and Patient, and the syntactic classes Subject and Direct Object. Goldin-Meadow *et al.* (2008) address this connection as well, albeit implicitly, when they discuss the correspondence between the orders they observed in their experiment and the dominant word orders in the languages of the world.

Goldin-Meadow *et al.* (2008) show that, for the events they tested, people used a consistent ordering for their gesturing: the Agent was typically gestured first, followed by the Patient, and followed by the Act. This order is consistent with the SOV word order, and it was found independently of the dominant order of the native language of the participants (which was not always SOV). Goldin-Meadow *et al.* (2008) conclude that the Actor-Patient-Act order is a natural order to mentally represent events.[5]

In the task described above, people are asked to communicate in a way they are not used to communicating. They can, however, use iconic gestures in order to describe the events, such as mimicking the shape of a glass when talking about a glass, or imitating a 'carrying' action. Thus, it can be said that the participants have a limited inventory of lexical items that they can use. When putting these lexical items together, participants *could* rely on the dominant word order of their native language. But Goldin-Meadow *et al.* (2008) show that in practice they do not do so. Participants do show systematicity in the ordering of the lexical items, but the order they choose is independent of their native language. The same behavior, some systematicity but not influenced by native language, is observed in the restricted linguistic systems discussed in chapter 2. Let us focus in more detail on the similarities between those systems and the experiment described here. The experiment puts participants in a situation that has the following properties:

1. Participants in the experiment cannot rely on their native language.
2. Participants in the experiment have a limited inventory of lexical items which they know will lead to communicative success.
3. The main goal of the participants when they produce gesture strings is making themselves understood, rather than adhering to an existing rule system.

The same properties hold for restricted linguistic systems:

- Unsupervised second language learners in the Basic Variety stage cannot rely on their source language, because their audience does not speak

will not consider these in the context of this dissertation.

[4]The difference in terminology most probably stems from the fact that the authors work in rather different disciplines.

[5]For a discussion of the nature of this conclusion, and alternatives to it, see chapter 5.

that language. They typically know some words already of the target language; they will put these words together as they see fit in order to make themselves understood, rather than paying attention to the precise grammatical rules of the target language (Klein and Perdue, 1997).

- In situations where pidgin languages emerge, speakers cannot use their native language, but there is usually already a limited lexicon of shared words. When using these words, the main goal is getting the message across, and not to adhere to an existing language.
- Children who develop Homesign do not have a native language at all when they are in this process. They have a limited inventory of largely iconic gestures that are understood by their family members. The homesigning child does not work towards mastering an existing sign language; the system is driven by the need for communication.
- The circumstances under which ABSL and NSL emerged are comparable, initially, to those under which homesigns emerge, but at a larger scale. In the Al Sayyid community, ABSL is used by both deaf and hearing people, but it is clear that none of the hearing people could count on their native language when communicating with the deaf members. Both systems emerged spontaneously, and speakers did not 'work towards' perfection of the system. In other words, the sign systems developed, driven mainly by communicative motivations.

Given these observations, a good way to describe the restricted linguistic systems 'in the wild' is the following: they all emerge in situations where speakers and hearers lack a common language: there is no shared set of conventions. Faced with this situation, the people involved start to improvise, using whatever means they have (iconic signs in the case of emerging sign systems; words from the target language in the case of Basic Variety speakers; words from the native language of pidgin speakers) and coming up with new strategies to put words together.[6] And this is exactly the situation that participants in the improvised communication task are faced with.

Thus, the improvised communication task puts people in circumstances that resemble those in which restricted linguistic systems arise. This makes the task a valuable source of evidence. However, the existing systems, despite their disadvantages, offer something that the lab task cannot offer: natural circumstances. For instance, the gesture sequences of a homesigning child all emerged spontaneously and were in no way guided by a lab task. For that reason, it is good to conceive of the improvised communication task as a way to *supplement* (rather than replace) the evidence from existing restricted linguistic systems. Moreover, in order to take advantage of the natural data from existing systems, as well as the amount of control that is possible in the lab, a valuable approach would be to work back and forth between existing restricted systems and the lab task. That is the approach that I will take in this dissertation.

[6] Of course, the individual circumstances of the systems will show differences, but, as we have seen in chapter 2, the systems show similarities.

Now that I have pointed out the similarities and differences between the improvised communication task and the restricted linguistic systems, it is time to look very carefully at the mechanisms that play a role in them, and what they can tell us about language evolution. That is what I will do in the next section: I will look at *bridge theories* (theories that describe the analogy between the situation of restricted linguistic systems and that of protolanguage speakers in our evolutionary history) that were provided in recent literature for restricted linguistic systems. I will compare and analyse them, and provide my own bridge theory for the improvised communication task. Subsequently, I will discuss empirical questions that can be studied using the improvised communication task.

4.3 Cognitive and communicative bridge theories

In chapter 2, we saw that Jackendoff's claim that speakers of restricted linguistic systems[7] apply fossil principles from protolanguage is intuitively appealing, but needs to be justified by pointing out why there is an analogy between situations in which restricted linguistic systems emerge, and protolanguage stages in the evolution of language.

In other words, we need to establish how exactly the analogy works between restricted linguistic systems and protolanguage. In Botha and de Swart (2009), several authors formulate initial steps towards a full specification of such an analogy. I will focus on the work of two authors: de Swart (2009) emphasises the role of cognitive structures and Roberge (2009) emphasises principles of communication. I will compare their views and show that they are different because of fundamental differences in the kind of data that they are based on. Moreover, I will show that a cognitive bridge theory necessarily also has communicative elements, and vice versa. This opens the possibility of formulating a bridge theory that combines both principles.

In what follows, I will refer to the views put forward by de Swart and Roberge as bridge theories, because this is the function that they are supposed to fulfil; I am aware of the fact that they would probably not be satisfactory bridge theories in the terms of (Botha, 2009b). Nonetheless, both views are valuable steps toward a full bridge theory.

4.3.1 De Swart (2009): cognitive structure

De Swart (2009) focuses on the expression of negation in the Basic Variety and observes that learners go through a stage where negative utterances are organised by topic/focus structure: they use negation preverbally, as an operator on

[7]Note that Jackendoff does not use the term 'restricted linguistic system'; I am using it here to refer to all phenomena (Basic Variety, pidgin/creole, newly emerging sign languages) discussed in chapter 2.

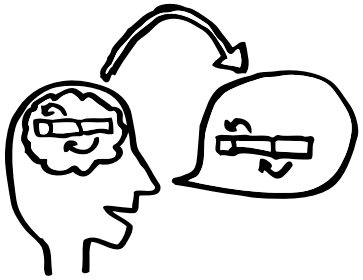


Figure 4.1: A cognitive bridge theory appeals to similarities between (1) cognitive structures present in speakers of restricted linguistic systems and (2) those in protolanguage speakers.

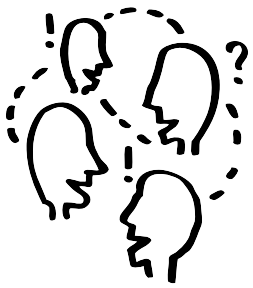


Figure 4.2: A communicative bridge theory appeals to similarities in the communicative situation between speakers of restricted linguistic systems and protolanguage speakers.

that part of the sentence that is in focus (generally the verb). They do this, even when preverbal negation is not part of the source or the target language.[8] On the basis of the data, de Swart (2009) hypothesises that the existence of negation as an operator on focus information was part of evolutionarily early language. She does this with an appeal to cognitive structure:

> I assume that the cognitive capacity for predicative structures [which is needed for the particular negative structure mentioned above - *MS*] is available in early humans before language develops. The pragmatically based combinatorics in the protolanguage stage reflects this pre-existing conceptual structure. (de Swart, 2009, p. 95)

Thus, de Swart (2009) stresses the importance of cognitive structures in justifying evolutionary conclusions from Basic Variety data. That is, cognitive structures in protolanguage speakers are compared to cognitive structures in L2 learners, and those structures are held responsible for the structuring of utterances in the BV as well as in protolanguage. A schematic depiction of this position is given in figure 4.1.

4.3.2 Roberge (2009): communicative principles

Roberge (2009) sketches a rather different bridge theory in an article about the creation of pidgins. He proposes to analyse pidgins as a form of language

[8] See the example of Swedish learners of French in (de Swart, 2009, p. 70–73).

creation, rather than a degraded form of language.[9] After showing several examples of creative innovations in pidgin languages (Roberge, 2009, p. 120–128), he proceeds to evolutionary conclusions we can draw from the structures observed in pidgin languages. A bridge theory, he proposes, should take communicative processes into account, by comparing the communicative situation of speakers of a pidgin language to that of our ancestral speakers of protolanguage.

Roberge pictures the situation of our evolutionary ancestors as follows. At some point in evolutionary history, proto-human society became increasingly complex (Johansson, 2005, p. 235), and small groups that had their own restricted communication patterns started to communicate among each other. The emergence of this cross-group communication led to a situation where linguistic features from the different 'group languages' competed with each other, and the ones with the greatest communicative efficacy would remain. The situation in which pidgin languages emerge can be described in a similar way: different groups using different languages are forced to communicate with each other, and they will use structures according to their communicative success. Thus, the analogy between the pidgin situation and that of our evolutionary ancestors should be about the communicative dynamics, and not about cognitive structures.

Roberge (2009) further hypothesises that the within group communicative systems of our hominid ancestors were rather limited, and that among group contact boosted the complexity of communication systems. This was the case because among-group communication resulted in a larger repertoire of meaningful elements, and this facilitated the creation of new combinations (Roberge, 2009, p. 131). In other words, according to Roberge, the situation of our evolutionary ancestors was as follows: when small groups of hominids with limited communicative systems started to interact with each other, this led to more available elements, and these elements were combined into more complex structures according to their communicative success. Pidgin languages can shed light on how this proceeded, because in situations in which pidgins emerge, there is a similar setup. A difference between the pidgin situation and the situation of our ancestors is that pidgin speakers have full modern languages at hand, but the fact is that they cannot use them when communicating. So the *process* in pidgin situations (creation of new linguistic features and competition among linguistic features for communicative success) is the same as that of our ancestors. A depiction of Roberge's view is provided in figure 4.2.

To wrap up, we have seen two bridge theories for restricted linguistic systems. De Swart (2009) claims that the structuring principles that can be seen in the Basic Variety are principles from cognition. She compares the cognitive structures present in speakers of the Basic Variety to those in our protolanguage speaking ancestors. Roberge (2009), on the other hand, sketches a bridge theory that compares the communicative situation of pidgin speakers and speakers of protolanguage. In other words, whereas de Swart focuses mainly on cognition,

[9]The latter is defended by Jackendoff; see chapter 2.

Roberge, with his focus on communicative success of competing constructions, describes protolanguage in terms of emerging *conventions*.

4.3.3 Cognition versus communication

The two bridge theories presented above apply to different systems: the Basic Variety in one case and pidgin languages in the other. To me, it does not seem surprising that a cognitive bridge theory was proposed for the Basic Variety and a communicative bridge theory for pidgin, because the circumstances under which they emerge are rather different. Speakers of pidgin are in a group of people speaking different languages and in order to communicate, they have to go around trying out lexical items from their native languages, as well as lexical items of the other languages that they may have picked up. Very soon in this process, a shared lexicon emerges, together with shared principles for organising utterances, and these together can be seen as an emerging system of new conventions. In other words, for pidgin speakers, a system of conventions emerges relatively early in the pidginisation process.

In the Basic Variety, on the other hand, speakers usually have as an audience speakers of the language they are learning. When they communicate, they can use lexical items of the target language and maybe some lexical items of their native language, if they don't know the right word. Moreover, they can come up with their own organising principles, but a crucial difference with the pidgin situation is that the audience of Basic Variety speakers will not copy any of the invented structures, because they are simply speakers of the target language and will apply the rules of this language.

To sum up, it is clear that a situation that gives rise to a pidgin language is more likely to develop shared conventions among speakers and hearers, and thus a bridge theory based on the emergence of conventions is more applicable to such situations. Speakers of the Basic Variety are more on their own in the sense that the hearers in the conversations they have will not share the structures they come up with. In other words, there is not really an emerging set of shared conventions, and the structures that come up are 'invented' by the speaker alone, and a bridge theory on the basis of cognitive principles seems more applicable here. For a summary of the differences between Basic Variety and pidgin formation, see table 4.1.[10]

Thus, pidgin is naturally paired with a bridge theory based on communicative principles, and the Basic Variety is naturally paired with a bridge theory based on cognitive principles. Does this mean that only one of the two is right? In other words, should we abandon one of the two bridge theories? I suspect we shouldn't. My hypothesis is that in protolanguage, both cognitive and communicative principles played a role. In the next section, I will explain how I

[10]In the table, I use the phrase 'formation of new conventions'. It might be confusing that I speak of the formation of new conventions in the Basic Variety situation. What I mean is the principles that govern the Basic Variety, which are, as we have seen in chapter 2, relatively independent of the source and target language.

Basic Variety	**Pidgin formation**
Only **speakers** have to improvise. Hearers will use the rules of their native language.	Both **speakers and hearers** have to improvise. None of them can use their native language.
Formation of new conventions is a **one way** process (initiated in the speaker).	Formation of new conventions is a **two way** process (speaker and hearer both participate in the process).

Table 4.1: The differences between (early) pidgin speakers and Basic Variety speakers.

envision this.

4.3.4 Cognition *and* communication

Above we have seen that there are two possibe answers to the question *what is responsible for the structure of utterances made in restricted linguistic systems (and, by inference, in protolanguage)*; namely, cognitive structures or communicative interaction. I propose that both cognition *and* communication are essential in processes where there is no fully established system of conventions, and individuals have to improvise in order to get their messages across:

The influence of cognitive structures: When a communicating individual constructs an utterance (1) with the help of a small established inventory of lexical items, but (2) in the absence of linguistic rules for combining these items into sentences, cognitive biases of this individual will influence the structure of the utterance.

The influence of communicative interaction: Different individuals may come up with different orders, but the constructions that are understood best by hearers will be established as linguistic conventions more easily.

But what to think of the two restricted linguistic systems discussed above: the Basic Variety and pidgin languages? Did we not notice that these two situations are essentially different, and give rise to different bridge theories? The two situations, I think, simply represent two extremes of a scale: in the Basic Variety, cognitive structures are much more influential than communicative interaction, and in pidgin formation, communicative pressures have a greater influence than cognitive structures of the individual speakers, just because of the different circumstances in which they emerge. Still, even de Swart and Roberge would not deny that in these systems both cognition and communication play a role. Let us have a look at how I imagine this.

In pidgin formation, cognitive structures might still play a role, besides the influence of communicative interaction, in the sense that these structures are

responsible for the utterance structures initially made by individual speakers. In other words, structures are either kept or abandoned according to their communicative success (as Roberge sketches the process), but the cognitive structures of the communicating individuals still have an influence on which structures are formed to begin with: individual speakers might have cognitive biases towards certain structures. Thus, the emphasis, in pidgin formation, might be on communicative interaction, but the decisions made by individuals (and thus their cognitive biases) still play a role.

On the other hand, something like convention formation plays a role in the Basic Variety. As described above, the communicative interaction between language learners and their conversation partners is different from that in pidgin, because language learners typically speak the target language with native speakers of the target language. These native speakers will not collaborate in establishing new linguistic conventions (like, for example, AgentFirst), because they already have the conventions (the grammar rules) of their language. Still, Basic Variety speakers might try out utterances with various structures and they will maintain the structures that are more successful in communication. In other words, also in the Basic Variety situation, communicative pressures might have an influence on the structures that are used.

And there is a second way in which communication might play a role in the Basic Variety. Think again about the principles according to which BV utterances are organised: do these principles stem directly from cognition? Are they simply mirror images of how these speakers organise their thoughts? They might be, but before we assume that, we should consider what that would mean: if the structure of utterances in the Basic Variety mirrors our thoughts, these thoughts should be organised in a language-like, linear way.

Do we organise our thought in some specific linear order? I am not a neuroscientist, and I cannot make any hard claims about the organisation of human thinking, but I do know that at least some of our knowledge is not organised linearly; for example, mental maps or mental images are not (Kosslyn, 1994). In other words, when speakers of the Basic Variety want to share a thought, they might have to go from something non-linear (image-like or otherwise) to something linear (an utterance). The former does not necessarily have an inherent order, and the latter does.

Thus, in the process of preparing a thought for communication, speakers create its structure. They go from a (potentially) non-linear thought to an utterance with linear organisation. This process is already the beginning of a communicative process. Of course, human thinking is influenced by language, and some of our thoughts might be language-like in structure, but I think we should be careful when ascribing a linear structure to all mental representations.

I realise that even when I put it as carefully as I just did, not everyone will agree with me that some of human thought has non-linear (for instance, image-like) properties. For example, Fodor (1975) describes mental images as 'pictures under description,' to 'avoid some obvious inadequacies of pictures'

(Aydede, 2010). The line of thinking I propose here can be seen as a mild speculation; I will get back to it in more detail in the next chapter, when I discuss the results of my experimental studies.

To summarise my view on the structure of thought and its role in communication, let me compare a normal communicative situation to one of a speaker of the Basic Variety. Generally, when we speak, we can only utter one word at a time. When we speak in a normal situation, using our native language, we simply use the order that is prescribed by the rules of our native language.[11] The situation for a speaker of the Basic Variety is different: there is no existing system of conventions for ordering elements of an utterance[12] and by imposing a certain word order a speaker can give extra information, because symbol positions can give information about semantic relations (Jackendoff, 2002, p. 238). One could say that these semantic principles for organising utterances stem from cognition, but cognitive representations do not necessarily have a linear order in the way utterances do. Thus, linearising mental representations into utterances already has a communicative flavour, because it prepares a thought for communication (to be understood by a hearer).[13] I thus propose that when people have to improvise, and put an utterance together without having a full system of conventions to fall back on, they might have cognitive biases that influence the utterance structures they choose. But it is not necessarily the case that the utterance structures that are found in restricted linguistic systems are direct mirror images of the mental representations of the speakers. I hypothesise that the formation of utterances in restricted linguistic systems is an active process that goes from multidimensional mental representations to linear structures.

When thinking about bridge theories, like I did above, the distinction between cognitive and communicative principles helps us draw evolutionary conclusions from restricted linguistic systems. I have analysed which situations cognitive and communicative principles apply to most naturally, but also how these principles can work together. This analysis will be put to use in the next section, where I will show that both cognition and communication play a role when drawing evolutionary conclusions from the lab data.

4.3.5 A bridge theory for improvised communication

Above, I discussed two bridge theories for restricted linguistic systems: one based on cognitive principles, the other on communicative principles. Moreover, I made the observation that in restricted linguistic systems, both cognition and communication play a role. Different situations (the situations I discussed were pidgin formation and the Basic Variety), however, have different characteristics

[11] Though it must be said that many languages allow free word order to some extent.

[12] Of course, there are the conventions of the target language, but as we have seen previously, speakers of the Basic Variety do not follow these conventions.

[13] The idea presented here shows rough similarities to Dan Slobin's notion 'thinking for speaking' (Slobin, 1996).

and in some, cognitive principles might play a greater role than communicative principles, and vice versa.

Let us now think about a bridge theory for the improvised communication task described in section 4.2. How can the gesture strings observed in this task tell us something about protolanguage? Which principles, cognitive or communicative, play a role, and how do they interact?

In the improvised communication experiment, there are no conventions concerning word order: people who are asked to convey meanings using gesture perform this task for the first time. Of course, they could revert to the ordering principles of their native language, but Goldin-Meadow *et al.* (2008) have shown that participants generally do not do that, because the same order, Actor-Patient-Act, is found in participants with different native languages. Goldin-Meadow *et al.* (2008) conclude from this that the order Actor-Patient-Act tells us something about the way in which events are represented mentally. If I would follow this line of thinking, I would end up formulating a cognitive bridge theory for improvised communication, and say that the structures found in improvised communication reflect mental structures, and that these mental structures are in turn comparable to the mental structures present in speakers of protolanguage. But I would like to suggest that a more dynamic analysis of the improvised communication task is possible.

Above, I suggested that even when we look at restricted language from a cognitive perspective, like we did for the Basic Variety, we can still see communicative principles at work. I suggested that the organisational principles observed in the Basic Variety are not purely mirror images of how thoughts are represented mentally, because not everything we represent mentally is represented in a linear way. The process of going from a thought to an utterance involves *linearisation* of the information: going from something that could be picture-like (a mental representation) to something in the shape of a string (an utterance).

This process of linearisation is, I will claim, very important in the improvised communication experiment. First, let me make clear why linearisation takes place in this setting. Participants in the experiment see a picture of an event on a screen, and are asked to convey the information in the picture using gesture and no speech. When they do that, they cannot represent all the elements in the picture simultaneously, but they have to make a sequence of gestures. Thus, they make a transition from picture to string.

Now, why is this essential? Because linearisation is (1) an active process, and (2) something inherently communicative. It is active, because every time a picture is shown, it has to be 'translated' into a string. It is essentially communicative, because the reason that the elements present in the picture have to be translated into a string is that they have to be communicated.

My claim that a process of linearisation is essential in the improvised communication task leads to a very specific prediction of the behaviour of participants carrying out this task. Contrary to what Goldin-Meadow *et al.* (2008)

claim, I predict that the Actor-Patient-Act order that they found is not necessarily the only order that will be found, because the linearisation process makes the participants' behaviour flexible: they can choose one order for certain situations, and a different one for others. This is possible because there are no strict, grammatical rules for order. Moreover, I predict that when they do change the order in their gesturing sequences, they will do this for communicative reasons. In other words, they change the order only when they think it makes a difference for the hearer (observer). Thus, they assume implicitly that different orderings lead to different interpretations. Below in section 4.3.6, I will specify how I will investigate these predictions experimentally, but first let me clarify how the improvised communication experiment connects to hypotheses about language evolution.

In the improvised communication task, people are asked to communicate in a way they are not familiar with. Using gesture and no speech, they are forced to improvise to convey the messages they are asked to convey; they have to start 'from scratch'. Previous research has shown that people in this situation will bypass the dominant order of their native language when sequencing lexical items in an utterance. What determines the ordering of their utterances? Above I have sketched situations where restricted linguistic systems emerge as situations in which new conventions are being formed. In the case of the improvised communication task, the formation of these new conventions is only just taking shape: there is only one 'speaker', who does not get much feedback about his communicative success. In other words, communicative pressures, in the way Roberge pictures them for the pidgin situation, cannot play a large role here, because there is not enough interaction for such pressures to have an effect.

What does play a role, clearly, are the cognitive biases of the participants in the experiment. But, as I have described above, I think there is more at stake than only a simple one to one relation between utterance structure and structure of mental representations of the participants in the experiment. I claim that when a participant in the experiment is presented with a stimulus (a picture of an event) and is asked to convey what is depicted, he or she will go through a process of linearisation of information: they go from a mental representation (which is not necessarily linear) to an utterance, in which only one element at a time can be expressed. My prediction is that communicating individuals, in the absence of stable linguistic conventions, will linearise information differently when order can be exploited to reflect differences in meaning. In chapter 5, we will see a specific situation where different stimuli ask for different communication strategies. More specifically, we will see an experiment in which two kinds of events with different semantic properties (extensional and intensional events) are communicated differently, using different orderings, and a second experiment in which it is shown that these different orderings lead to different interpretations. These results suggest that from the early stages of communication onwards, speakers and hearers are sensitive to meaningful differences conveyed by word order. In chapter 6, we will see how participants

organise their improvised gesture strings when we go beyond simple events and add information about the time of an event.

To summarise, in the improvised communication task, we simulate a situation in which there are hardly any existing linguistic conventions. The behaviour we observe from participants tells us something about their cognitive biases, and their intuitions about successful communication. The fact that, as has been shown by Goldin-Meadow *et al.* (2008), participants by-pass the dominant order of their native language shows that participants in the experiment do not simply fall back on the linguistic rules of their native language in order to communicate, but that they use intuitions and biases that are more fundamental than that.

The key to my bridge theory is the similarity of the situation individuals (in the experiment and in prehistoric times, respectively) find themselves in: they have to linearise information into a string, in the absence of a stable system of conventions. An obvious difference between us and our protolanguage speaking ancestors is that we have language, and this, of course, influences our behaviour. Having language must have had an influence on our cognitive capacities, and for this reason, some people claim that we cannot compare behaviour observed in modern people to that of our evolutionary ancestors: our minds might work differently from those of our ancestors. The fact that people in the experiment bypass the rules of their native language suggests that we should not be too worried about this:

> [T]he ordering found in nonverbal tasks appears to be more robust than the ordering found in language; speakers of four different languages used different orders in their spoken sentences, yet all displayed the same order on two different nonverbal tasks. (Goldin-Meadow *et al.*, 2008, p. 9167)

What exactly is the status of the orderings found when people bypass their grammar rules? Of course, I cannot make hard claims about this, but it could be that in human minds, evolutionarily older and newer systems exist simultaneously, and the improvisation task triggers people to use 'old' mechanisms. Ideas parallel to this (that older and newer mechanisms in behaviour or cognition coexist) are put forward in various disciplines. In computational modeling, for example, researchers use incremental evolution in order to teach complex behaviour to neural networks (Gomez and Miikkulainen, 1997). In psychology, cultural evolution is characterised as cumulative; Tomasello (2001) calls this the ratcheting effect. And language itself has been characterised, by Hurford (2011), as a layered phenomenon, exhibiting old and new structures:

> In each language, we find vestigial one-word expressions and protosyntactic (2-, 3-word constructions) keeping company with more fully elaborate syntax. All languages have the possibility of conveying propositional information without the benefit of syntax. English

> speakers use a single word, Yes or No and pragmatic inference identifies the particular proposition which is being confirmed or denied. No language lacks such words. They are a part of any language — just not of interest to syntacticians. (Hurford, 2011, chapter 5)

These examples do not provide conclusive evidence that human linguistic behaviour is a cumulative phenomenon, but it is interesting that similar ideas have been put forward in different contexts, and it is certainly a worthwhile direction for further research.

To conclude, I will not claim that the rules of the languages of the world result solely from the intuitions and biases that play a role in the linearisation of information (this would be a very odd claim, because there is such a great variety of languages and linguistic rules), but the improvised communication experiment does catch an important aspect of what plays a role when a linguistic system emerges. Possibly, even, an evolutionarily old system is at play (either a cognitive or a cultural system).

4.3.6 From a bridge theory to research questions

Above I have presented a way to collect restricted linguistic data in the lab, and I have formulated the principles that allow us to draw evolutionary conclusions from the observations made in the lab. Now it is time to formulate specific questions that we can investigate using the experimental setup described here.

A good place to start is the description of very simple events (simple propositions) in modern language, which are described by a verb and its arguments, like the following examples:

(1) The tree hits Fred.

(2) Fred eats an apple.

In these examples, situations are described in which X does something to Y. For both sentences, we understand who did what to whom, because we know the rules of English, and English has a fixed word order. It is interesting to see how this kind of event is expressed in simple language systems, because utterances of this type in protolanguage might have been the basis of dominant word orders in modern languages. Thus, by looking at how simple propositions are expressed in improvised communication, we can, ultimately, draw conclusions about the origins of word order. In the next section, I will specify the research question I will work with, and in chapter 5, I will describe two experiments that investigate these questions. In the experiment, I will extend the claim put forward by Goldin-Meadow *et al.* (2008), that SOV is the basic word order.

A logical next step, after looking at simple propositions, is to look at what happens if we go one step further: what if we make utterances more complex by adding information to a simple proposition? Simple propositions, in a language system where there is no verb inflection, are very much bound to the *here and*

now: there is no obvious way to specify that an event happened at some other time. But we have already seen in chapter 3 that thinking and communicating about things that are not here and now is a very central human capability that might have been a key step in the emergence of language. One would expect this property to show up in restricted linguistic systems. Consequently, the question is: how is the information that an event happened in the past or the future expressed in restricted linguistic systems? In section 4.5, we will first have a look at mechanisms from existing restricted linguistic systems, and on the basis of that, I will formulate research questions that will be investigated in chapter 6.

4.4 Expressing simple propositions: the origins of word order

In chapter 3 we have observed that many sentences in natural language express simple propositions, and these simple propositions are expressed by a verb and its arguments. These simple propositions are an interesting place to start looking at the emergence of linguistic conventions, because the way in which simple propositions are expressed in a language has a strong connection with the dominant word order in that language.

Languages with a fixed word order have a conventionalised way to present the verb and its arguments: each element has a fixed place. (Languages that do not have a fixed word order, might use other means to make clear who plays what role in a described event, for example with the help of case marking, but the majority of the languages of the world have a (relatively) fixed word order (Dryer, 2011).) This has advantages over just concatenating words without a conventional order:

> With just symbol concatenation, *eat apple Fred* and *eat Fred apple* might be used to convey exactly the same message. In this particular case there would be no problem, because of the pragmatics of the words involved. But in *hit tree Fred*, did Fred hit the tree or did the tree hit Fred? Though the larger context might tell us, the pragmatics of the words alone do not. (Jackendoff, 2002, p. 247)

Whereas full languages typically have a conventional way to order the elements of sentences, restricted linguistic systems do not have hard rules about the order of the elements in utterances, although some semi-systematic patterns have been observed. The question we can now ask is the following:

> *How is the trajectory from language systems with no organisational principles to full language with fixed word order? Which processes (cognitive or communicative) played a role in establishing these word orders?*

In order to find an answer to this question, we will first look at word order in the languages of the world, subsequently at the dominant word orders found in restricted linguistic systems, and finally, we will investigate word order in the improvised communication task in the lab, in chapter 5.

4.4.1 Word order in the languages of the world

There is a great deal of variety in the languages of the world. In the case of basic word order, the preferred order of Subject, (direct) Object and Verb in transitive sentences like 'the dog chases the cat', all possible orders are represented: SOV, SVO, VSO, VOS, OSV and OVS (Dryer, 2011). Not all orders occur equally often, though. Orders that start with O are quite rare, and orders starting with S are the most common. The types SOV and SVO together make up about 76% of the totality of languages (Dryer, 2011). Not surprisingly, these two word orders have received a lot of attention. Both Newmeyer (2000) and Gell-Mann and Ruhlen (2011) claim that protolanguage had a dominant SOV order. Langus and Nespor (2010) give an overview of historical and comparative linguistic sources reporting that the SVO order is the preferred structure for syntax.

In what follows, we will see that the two types SOV and SVO emerge as the most common word orders in restricted linguistic systems.

4.4.2 Word order in restricted linguistic systems

In chapter 2 we have seen an overview of restricted linguistic systems that described the circumstances in which these systems appear, as well as some of their organising principles. Now we will have a look, specifically at the dominant word orders in these systems, in order to see if these restricted systems can provide conclusive evidence about the emergence of word order in modern languages. Before I start, note that to describe the orders in restricted linguistic systems, I use 'S', 'V', and 'O' instead of 'Actor,' 'Patient' and 'Act,' the terms used above to describe the elements in improvised communication. I realise that my usage of S,V and O is potentially problematic, exactly because of my claim that restricted systems are governed by semantic principles and not by syntactic rules. On the other hand, S, V and O are used elsewhere (e.g., Hurford (2011)) to describe word order in restricted linguistic systems, and even to describe gesture sequences in improvised communication (Langus and Nespor, 2010).

In the **Basic Variety**, the following structure is observed: NP_1-V-(NP_2) (Perdue, 1993). This is consistent with the word order SVO. As noted in chapter 2, the fact that all target languages are SVO languages, could be responsible for the SVO ordering found in the Basic Variety (Hurford, 2011). In chapter 2, it was shown that speakers sometimes deviate from SVO word order. Interestingly, this is done in specific situations, as is shown in the example on page 37: 'stealing bread girl,' uttered by a Punjab learner of English in a situation

System	dominant order	modality
Basic Variety	SVO	spoken
Pidgin/creole	SVO	spoken
homesign	(S)OV	signed
ABSL	(S)OV	signed
NSL	SOV and OSV	signed

Table 4.2: Dominant word orders in restricted linguistic systems.

where 'girl' is focus information. This could indicate that a semantic principle like FocusLast applies here (Hurford, 2011, chapter 5.7).

Pidgins and creoles are reported to be mainly SVO order (Hurford, 2011), although some Niger-Congo creoles are SOV (Holm, 1988). The orders found for pidgin and creole might be influenced by their superstrate and substrate languages.

In **homesign**, the Subject is very often not expressed, and the remaining Object and Verb are expressed in OV order (Goldin-Meadow *et al.*, 2009).

In **Al Sayyid Bedouin Sign Language** (ABSL), like in homesign, the Subject is relatively often not expressed. Sandler *et al.* (2005) report that the dominant word order is again OV, although other orders are found as well. Whenever the Subject is expressed, this results in SOV order. As described in chapter 2, verbs are sometimes split into two separate verbs, so that for a sentence like 'man pushes woman,' a sign sequence 'MAN PUSH - WOMAN FALL' is created; thus, an SV-OV pattern is used here.

In **Nicaraguan Sign Language** (NSL), there is also a general preference to put the verb in final position. Many word orders are found though; Senghas *et al.* (1997) report, for example, that for transitive verbs with an animate Subject and an inanimate Object, the orders OSV and SOV are both used (p. 555). They contrast this with transitive verbs with two animate arguments (such as 'a man pushes a woman'), in which the verb is split into two separate verbs, *and* the Subject is always signed in the first part, and the Object in the second ('MAN PUSH - WOMAN FALL').

From the observations above, we conclude that the two word order types SVO and SOV, the most common among the languages of the world, play an important role in restricted linguistic systems as well. Modality seems to have an influence as well (see table 4.2). Spoken restricted systems (Basic Variety and pidgin/creole) seem more likely to have SVO order than manual systems (homesign, ABSL, NSL), which tend to have mainly (S)OV order. Could it be that SVO order cannot exist in manually signed language systems? This would certainly seem likely given the data considered here, but in chapter 5, I will show that this is not the case.

The data above also show that word order is not so strict in restricted systems, and in the next section I will discuss the possibility that the way in which speakers vary their word order is evolutionarily interesting.

4.4.3 Variation in word order

In many of the restricted linguistic systems discussed above, something like a dominant word order can be defined, but other orders are used besides the dominant word order. This is not surprising when we think about it. Restricted linguistic systems are systems in which linguistic conventions are still in development; there are no hard rules yet. We have also seen that the cases where speakers deviate from the dominant word order have interesting semantic properties. In NSL, the subject is only put in front position when it is important to convey the information that the subject is indeed the agent. In situations where the direct object is inanimate, it will be immediately clear that the subject is the agent, and in these situations it is not a problem to use OSV word order. In the Basic Variety, speakers sometimes deviate from the dominant SVO order, in cases where the agent is clearly focus information.

Thus, in restricted linguistic systems, one can speak of dominant word orders, but it is more interesting to look at those situations where speakers vary their word order. This may sometimes be caused by circumstantial, uninteresting factors, like the source language of the speaker in the case of SOV strings used by speakers of the Basic Variety. In other cases where word order deviates from the dominant word order, there are interesting semantic properties at stake, and we see reflections of the semantic/pragmatic principles described by Jackendoff.

In my opinion, restricted linguistic systems should be seen as situations in which there are emerging conventions, which are partly responsible for the word orders that can be observed in them. On the other hand, there is variation in word order by individual speakers, and this variation occurs in specific semantic or pragmatic circumstances. The structures that emerge in these situations are the result of cognitive biases of the speaker, *plus* the way the speaker linearises a mental representation into a string (cf. section 4.3.4) in order to communicate successfully.

In chapter 5, we will start from this observation and look at variation in word order by participants in the improvised communication experiment. The experimental setting allows us to create very precise stimuli, with which specific semantic or pragmatic circumstances can be investigated. We will use this ability by studying the influence of the distinction between *extensional* and *intensional* verbs, a distinction described and analysed in semantics and philosophy of language, on the structuring of utterances.

4.5 Adding information to a proposition: temporal displacement

The second topic I will focus on by means of an experimental study has to do with complexification of simple utterances. What if a speaker wants to convey a message that has more information than only a description of an action and

its participants? An example of such a more complex message is one in which not only an event is described, but it is specified when the event took (or will take) place.

In chapter 3 it was already pointed out that a crucial feature of human cognition and language is that we can think and talk about things that are not here and now. In other words, we can think and exchange information about things that are remote in space and time.

Full languages have sophisticated means to express information about events that are remote in time. Early forms of language most likely did not have the means to do this. But if talking about temporally remote things is such a crucial feature of language, we can ask ourselves:

> *Did our evolutionary ancestors express information about the past and the future? If they did, how did they do it in a linguistic system that is simpler than modern languages? Which processes (cognitive and/or communicative) play a role in temporal displacement?*

In the remainder of this section, I will first focus on temporal information in full languages, then look at the expression of temporal information in restricted linguistic systems, in section 4.5. Subsequently, we will investigate how temporal information is expressed in the improvised communication task in the lab, in chapter 6.

4.5.1 Temporal information in modern languages

There is a large body of literature about the expression of temporal information in modern languages, as modern languages have sophisticated means to express temporal information, and there are large differences between different languages as to how temporal information is expressed. The following is only a very crude summary, just enough to give the reader an idea of what kinds of temporal information can be expressed in modern languages.

The many existing grammatical features in modern languages that have to do with time are used to express information about *tense* and *aspect*. Tense expresses at which point in time an event took place or will take place. Aspect expresses the speaker's *viewpoint* on the event: is the event finished, or still going on? In many languages, tense and aspect are expressed through inflection on the verb. Here are some simple examples in English: 'I will cycle to the office' refers to the future; 'I cycled to the office' refers to the past; 'I was cycling to the office' refers to the past too, but takes a different viewpoint, by emphasising the 'ongoingness' of the situation. I will now give a brief overview of how tense and aspect are described systematically in linguistics.

Reichenbach (1947) describes the possibilities of tense and aspect using three notions, 'point of the event' (E), 'point of speech' (S) and 'point of reference' (R), which links E and S. These points are situated on the time axis, and the verbal tenses of languages can be expressed in terms of the positions

of the three points with respect to each other.[14] The relation between R and S tells us whether we focus on something in the past, the present or the future: when past tense is used, R comes before S; when present tense is used, S and R coincide; when future tense is used, R comes after S. In other words, tense specifies the reference point R on the time axis, with respect to the point of speech.

The relation between R and E says something about the viewpoint on the event: R can coincide with E, or come before or after it. For example, we can model the difference between *I saw John* and *I had seen John* with the help of the relation between R and E. For the simple past form *I saw John*, R comes before S (because it is past tense), and R coincides with E:

(3) I saw John.
E,R — S

For the past perfect form *I had seen John*, R comes before S, but R comes after E: we 'look at' the event from a point where the event is already in the past:

(4) I had seen John.
E—R—S

The tense forms of modern languages can be described using Reichenbach's E, R and S. But verb forms are not the only means to express temporal information in languages. *Temporal adverbs* can be used to express both temporal and aspectual information. We will now have a look at adverbs of temporal location, adverbs that specify event times; these adverbs also play a role in restricted linguistic systems (see the next section).

Adverbs of temporal location are used to indicate where on the time axis an event that is described takes place. Thus, whereas a sentence like *I saw John* gives us the information that the event (me seeing John) took place somewhere in the past, in the sentence *I saw John yesterday*, the time of the event is specified more precisely. But one should note that it is not E (the time of the event) directly that is specified by a temporal adverb, but rather the reference point R. This can be seen from the example *I had seen John yesterday.* In this example, 'yesterday' does not refer to the point in time where I saw John, but to R, which lies *after* E. In other words, when I say *I had seen John yesterday*, it can be the case that I met John the day before yesterday (Reichenbach, 1947).

To summarise, verb forms in languages specify the relations between R and E and between R and S. Temporal adverbs determine the time of R. In a simple past sentence (*I saw John yesterday*), 'yesterday' specifies R, and thereby, indirectly, the time of the event (E), because these two coincide. In a

[14]Reichenbach (1947) applies his system to English; it has been shown that to model the tense forms of e.g. French, one needs more than Reichenbach's E, R and S (Verkuyl et. al., 2005).

past perfect sentence (*I had seen John yesterday*), 'yesterday' specifies R, and the verb form tells us that the time of the event comes before R.

De Swart (1999) makes Reichenbach's analysis of temporal adverbs more precise by taking topic and focus structure into account. As pointed out in chapter 3, the placement of a temporal adverb expresses topic-focus structure. Compare the following two sentences, from de Swart and de Hoop (2000):

(5) Jane left at six o'clock.

(6) At six o'clock, Jane left.

Sentence (5) is a typical answer to the question 'when did Jane leave?', whereas sentence (6) is a typical answer to the question 'what happened at six o'clock?'. Note that this is not a hard rule: when the intonation is adapted, sentence (5) can also be an answer to the question 'what happened at six o'clock?'. What is crucial about these two constructions is that sentences in which the temporal adverb is fronted are more suitable to introduce a *temporal frame* for the main clause (de Swart, 1999). A temporal frame is a given time on the time axis, in the context of which the rest of the utterance can be interpreted. Or, put in Reichenbach's terms, it specifies the temporal location of R.

In the next section, we will see that the data suggest that restricted linguistic systems use this mechanism (introducing a temporal frame by using a fronted temporal adverb) in order to specify the time at which an event takes place.

4.5.2 Restricted linguistic systems: the displacement strategy

As we have seen above, most modern languages use inflection on the verb to express temporal information. In restricted linguistic systems, however, verbs are generally not inflected, or at least not in a systematic way. This means that the standard way to refer to past and future is not available in restricted linguistic systems. Despite this, as we will see below, past and future are referred to in restricted linguistic systems. Speakers of restricted linguistic systems can do this by employing temporal adverbials. This means that, speaking in terms of Reichenbach, they are able to specify a reference point R on the time axis. In other words, when they describe an event in combination with a temporal adverb, R is specified. But they do not have the grammatical means to distinguish between cases where E coincides with R and cases where E comes before R, because they lack the verb forms to express this distinction. This is not necessarily problematic, because R can simply be assumed to coincide with E. The temporal adverb is then used to specify the time at which an event took place.

Let us have a look at some examples. Benazzo (2009) discusses the expression of temporal information in the Prebasic Variety. This is a stage in the learning process of second language learners who learn a language outside the classroom, and it precedes the Basic Variety stage. In the Prebasic Variety,

speakers do not use many verbs, and utterances are generally organised around noun phrases. Utterances typically consist of target-language-like nouns, adverbs and adjectives. Thus, in the Prebasic Variety, the inventory of speakers is very limited, even compared to that of Basic Variety speakers. Still, Prebasic Variety speakers refer to events in the past and in the future.

The following example shows utterances of a Spanish learner of French (Benazzo, 2009, p. 28):

(7) [interviewer asks: what about your husband?]
en* el* hopital
(Spanish: in the) hospital

(8) demain ++ permis
tomorrow [pause] permit (=to leave the hospital)

(9) et lundi à l'hopital
and Monday to the hospital

The utterance in (7) is about the current situation, so no event time is specified (the event time then follows the time that was talked about previously in the conversation, which is in this case 'now'). The utterances in (8) and (9) are about other points in time, and there, the expressions 'tomorrow' and 'Monday' are used to introduce a temporal frame for the events that are described ('permit' and 'to the hospital' respectively). As indicated above, the events are taken to coincide with the temporal frame that was introduced. In other words, 'permit' (to leave the hospital) will take place tomorrow, and 'to the hospital' will take place on Monday.[15]

Thus, in the Prebasic Variety, temporal adverbs are fronted in order to specify the time of an event, when an event does not take place now. Similar constructions are found in the Basic Variety (Benazzo, 2009, p. 29):

(10) [interviewer asks: what are you doing here? Are you working?]
avant je [travaj] / maintenant non
before I work / now no

Again, the event time is specified with help of adverbs in the initial position. The same behaviour is observed in homesign (Benazzo, 2009, p. 40):

(11) [someone talks about his working situation:]
before [pointing over shoulder] GOOD
now [pointing downwards] SO SO

Thus, reference to past and future is made by marking an event time with an adverb, and this information is put in the sentence-initial position. In other words, in restricted linguistic systems, adverbs are used to mark a temporal

[15]Because of the expressive limitations of speakers in the Prebasic Variety, their utterances are context dependent, and depend heavily on pragmatic inferences.

frame, and thereby they displace the event that is talked about from the here and now. I will call this strategy *the displacement strategy.*

Full languages apply this strategy as well, but because inflection of the verb is possible in full language, it is possible to specify more fine grained information, such as the distinction between *I saw John yesterday* and *I had seen John yesterday.* In these examples, R and E are specified separately. Restricted linguistic systems cannot do that: they can only specify an R, and then it is implicitly assumed that E takes place at R. In other words, restricted linguistic systems are less expressive than full languages because much of the information is left implicit. But this does not mean that in restricted linguistic systems, reference to the past and the future is impossible, for the displacement strategy can be applied in order to do that.

In chapter 6, we will investigate whether the displacement strategy is applied when people are asked to communicate about the past and the future in the improvised communication experiment. It would be striking if participants in the lab would apply the same strategy as the speakers and signers of the restricted systems mentioned here.

In this way, the improvised communication experiment is applied to verify a structure observed in other restricted linguistic systems. But on top of that, the experiment makes it possible to look in further detail at the interplay between the displacement strategy and the way simple propositions are described: does the displacement strategy influence the way in which simple propositions are expressed? This question will be investigated in chapter 6 as well.

4.6 Conclusion

In this chapter I have described a way in which we can collect more restricted linguistic data: by creating a lab situation in which participants have to improvise in order to get their messages across: the improvised communication experiment. This experimental method offers us a way to collect data in a more controlled manner than was possible before.

I proposed that we can draw conclusions about the structure of protolanguage from the strings observed in the improvised communication experiment. In the absence of a stable system of conventions, participants in the experiment have to linearise information into a string, just like our protolanguage speaking evolutionary ancestors had to do.

Subsequently, I proposed two research topics that we can investigate using the improvised communication experiment. The first concerns the expression of simple propositions: observing how an action and the participants in an action are talked about in improvised communication can eventually help us to find out more about the emergence of dominant word orders in modern languages. The second concerns the expression of temporal information, which is a way to add information to a simple proposition. It will give us insight into the way in which utterances that express simple propositions can be extended into more

complex ones.

The next two chapters will be empirical in nature: they present the lab experiments that I have conducted to investigate the two research topics introduced here.

CHAPTER 5

Improvised communication and the semantic origins of word order[1]

5.1 Introduction

In chapter 1 of this thesis we have seen that, once we acknowledge that language is not a monolithic whole and emerged gradually, the investigation of protolanguage as a hypothetical stage in the emergence of language is a logical step. In chapter 2, it was shown that one way to investigate the structure and function of protolanguage is by looking at restricted linguistic systems. Analysis of data from various restricted linguistic systems suggests that these systems are governed by semantic (and pragmatic) principles. These principles might therefore have played a role in protolanguage. This step from restricted language to hypotheses about the emergence of language, as we have seen, is not completely unproblematic. One remaining issue is the fact that data collection is labour intensive and that the data is not collected in a controlled environment. Another issue with existing restricted linguistic systems is the kind of inference we make from the data to the evolutionary conclusions: what exactly is the nature of this kind of reasoning? How can we justify it? In the existing literature, this is still underdeveloped.

A close investigation of the discipline of semantics in chapter 3 showed that existing characterisations of meaning fall into three categories, and each approach stresses the importance of one aspect of meaning: meaning as the

[1] Part of the research presented in this chapter was originally published as Schouwstra *et al.* (2011).

relation between language and cognition, meaning as the relation between language and the world, and meaning as a conventional phenomenon.

A characterisation of meaning in cognitive terms is the most straightforward account when meaning is investigated in an evolutionary context. It is, however, important to acknowledge that the other two views of meaning are important when the emergence of language is discussed; this becomes clear when issues like compositionality and grammaticalisation are discussed.

It was shown in chapter 4 that as far as restricted linguistic systems are concerned, the conventional nature of language is pushed to the background somewhat, because language users in these situations typically depend on improvisation and there is only the beginning of a system of newly formed conventions. In such situations, individual decisions about how to communicate are thus very important. It should be taken into account, however, that when people construct utterances in a restricted system, this does not simply reflect the structure of their cognitive processes: when they construct utterances, they linearise information into a string.

The goal of this chapter is, first of all, to describe and extend an experimental approach put forward in Goldin-Meadow *et al.* (2008), which forces people to communicate about simple situations in an improvised manner: by using gesture instead of speech. I will argue that this kind of experimental approach can be used to supplement and extend the evidence that is found in the restricted linguistic systems described in chapter 2.

The particular study presented in this chapter concerns the emergence of word order. It has been suggested in the literature (see, e.g., Newmeyer (2000)) that the word order of protolanguage was SOV. A similar scenario is suggested in Goldin-Meadow *et al.* (2008). In this chapter, additional complexity is introduced to this issue and it is shown that semantic properties have an influence on utterance construction in improvised communication, and may therefore have had an influence on word order in evolutionarily early languages.

5.2 Background

5.2.1 Goldin-Meadow *et al.* (2008)

Goldin-Meadow *et al.* (2008) investigate how people sequence information when they are asked to communicate about simple events using gesture and no speech. Forty adults (speakers of 4 different languages: English, Chinese, Spanish and Turkish) were asked to describe simple events depicted in vignettes using only gesture and no speech. Each vignette depicted a motion event: a simple event containing an actor, patient and some act, in which the act involved motion.[2] Before they carried out the gesturing task, the participants

[2]These events are typically described by transitive sentences; the study also contained intransitive actions, but those are not taken into account here.

were asked to describe each vignette in a simple declarative sentence in their native language.

A striking difference was found between the order in these sentences and those in the gestured sequences. The different semantic elements in the vignettes were put in different orders when described in a spoken sentence, depending on the language of the participant. E.g., English is an SVO language, and the vignettes were described in Actor-Act-Patient order when put in a sentence. Turkish, on the other hand, is an SOV language and the typical order of the semantic elements of the vignettes was thus Actor-Patient-Act.[3] In other words, when using normally spoken language, participants typically followed the order of their native language. When they were asked to act out the vignettes without using speech, however, they did not follow the order of their native language. Instead, all participants used the same order in this task: Actor-Patient-Act.

Goldin-Meadow *et al.* (2008) point out that this order, Actor-Patient-Act, corresponds with the sentence order SOV, which is one of the two most common word orders in the languages of the world. Moreover, they present a non-communicative task. People were given a set of 3 transparant pictures, each portraying one part of the event (so, separate transparancies were made for the actor, the patient and the act). In order to complete the picture of the event, the transparancies had to be stacked. No instructions were given about the order in which they had to be stacked, but people showed consistency in their stacking orders: there is a general preference for Actor-Patient-Act order in this setup as well.

Goldin-Meadow *et al.* (2008) conclude from this that Actor-Patient-Act (SOV) may reflect a natural sequencing for representing events. They speculate that this order is not an outgrowth of communicative efficiency but rather the way in which events are represented mentally in individuals, and they refer to the results of the non-communicative task to support this claim.

5.2.2 Langus and Nespor (2010)

In Langus and Nespor (2010), the starting point is a view of the human faculty of language as a modular system, in which different cognitive systems are responsible for different tasks: the conceptual system provides the meaning of linguistic utterances, the sensory-motor system produces and perceives the actual sounds and signs of language, and the computational system of grammar links meaning with sounds by generating the structure of sentences.

Langus and Nespor contrast the conclusion drawn by Goldin-Meadow et al. that there is "a strong predisposition for the SOV order in simple improvised communication," with the observation that 'convergent evidence in theoretical linguistics points to the universality of SVO as the basic word order for the

[3]See the previous chapter for more information on the dominant word orders in the languages of the world.

computational system of grammar' (Langus and Nespor, 2010, p. 3). They note that the same dichotomy between SVO and SOV can be seen in two cases of atypical language acquisition: in creoles, SVO order is observed often, and in homesign, the object-verb order is generally seen.

Langus and Nespor hypothesise that an SOV (Actor-Patient-Act) order in the gesturing of simple events is the order that is preferred whenever the computational system of grammar is by-passed, and results from a direct interaction between the sensory-motor system and the conceptual system. Only in cases where people need their computational module, Langus and Nespor suggest, would people prefer SVO order.

They present four experiments to support these claims. In the first experiment, speakers of Italian and speakers of Turkish were used, the setup was similar to that of Goldin-Meadows experiment, and the results from Goldin-Meadow *et al.* (2008) were replicated.

In the second experiment, the improvised communication of complex events was investigated. For this experiment, vignettes were used that represent complex situations, such as, e.g., *[the man tells the child [that the girl catches a fish]]*. It was found that both Turkish and Italian participants gesture the subordinate clause after the main clause, an order typical of SVO languages. It is concluded that, apparently, the SOV order in improvised situations does not generalise to more complex SOV language-like constructions and that participants thus did not use their computational system of grammar to produce the strings.[4] Moreover, they concluded from the results of experiment 1 and 2 together, that all participants (the Italian participants in experiment 1, and the Turkish participants in experiment 2) bypass the structure of their native language when communicating in this improvised gesturing setting.

In the third experiment it was shown that in gesture *comprehension*, SOV order is preferred as well: it led to the shortest reaction times in a gesture comprehension task carried out with Turkish (SOV) and Italian (SVO) participants. Thus, the word order of one's native language is bypassed in both the production and interpretation of improvised communication, and SOV ordering is preferred instead.

In the fourth experiment, they investigated speech comprehension, using, again, speakers of Turkish and Italian. In order to test this, they created prosodically flat strings of words in the native language of the participants. The strings corresponded to the scenes in the vignettes and they were put in all possible orders of Subject, Object and Verb (SOV, SVO, OSV, OVS, VSO and VOS). After hearing each string, participants were asked to choose between two vignettes, one of which corresponded with the speech string. The reaction times of the participants per item were analysed.

As expected, the participants showed a preference for speech strings in the order of their native language, i.e., Turkish participants were fastest with SOV

[4]This conclusion might seem puzzling. We will get back to it later in this chapter, in section 5.7.4.

strings and Italian participants were fastest with SVO strings. Looking at all six possible orders together, however, revealed consistent preferences among Turkish and Italian participants. For Italians, all orders in which the verb occurred before the object (SVO, VOS and VSO) resulted in faster reaction times than orders in which the verb occurred after the object (SOV, OVS and OSV). This was true for the Turkish participants as well, despite their good performance on SOV order. Langus and Nespor conclude that this is evidence showing that the computational system of grammar privileges the Verb-Object orders (Langus and Nespor, 2010, p. 17).

Closer analysis of the two groups of word orders, however, suggests a different explanation: the 'slow' group of word orders, consisting of SOV, OVS and OSV, contains two orders where O comes before S (OVS and OSV), whereas the other, faster, group contains only one order where this is the case (VOS). Greenberg (1963), shows that word orders where O precedes S are very rare among the languages of the world and it is suggested that this is because O<S is hard to comprehend (van Leeuwen, 2010). Thus, the occurrence of O<S in two of the three word orders in the slower group offers an alternative explanation of the findings, and the results cannot be seen as support for Langus and Nespor's claim that orders in which V comes before O are faster.

5.2.3 The origins of word order

The authors of the articles discussed above agree on one thing: they claim that SOV order, found in the gesturing sequences for motion events is in some way a reflection of conceptual structure, i.e., the structure of mental representations:

> Our data suggest that the ordering we use when representing events in a nonverbal format is not highly susceptible to language's influence. Rather, there appears to be a natural order that humans (regardless of the language they speak) use when asked to represent events nonverbally. (Goldin-Meadow *et al.*, 2008, p.9167)

Another conclusion that is drawn in both publications is that the results from these improvised communication experiments are important in the discussion about the emergence of language.

Goldin-Meadow et. al. suggest, as indicated above, that humans represent events in SOV order, and that this order is independent of the language we speak. This order, they suggest, might have been important in the early stages of the emergence of language:

> [T]he ordering seen in our nonverbal tasks may shape language in its emerging stages. (Goldin-Meadow *et al.*, 2008, p. 9167)

Langus and Nespor draw a similar conclusion. They claim that simple communication prefers SOV order and they connect this to the emergence of language:

> Our results suggest that also our linguistic abilities coexist with, and possibly derive from, a more primitive form of communication that relies on the direct mapping between the conceptual and the sensory-motor system. (Langus and Nespor, 2010, p. 21)

5.2.4 Communication, or mental representation?

Goldin-Meadow *et al.* (2008) claim that the ArPA order found in the participants' gesturing cannot come from communicative pressure, because participants used the same order when they were asked to do a non-communicative task (the transparancy task described above). For this reason, they speculate that, as mentioned above, ArPA may reflect a natural sequencing for representing events.

But what does it mean for semantic elements to be sequenced? As I observed in chapter 4, semantic elements that occur together in an event can be mentally represented in many ways, some of which do not have any inherent order (for example in a mental map or a mental image). So, linearisation of these elements only takes place when we are forced to impose an order on the semantic elements.[5] A situation that typically forces us to linearise information is communication: when we communicate about events (using either speech or gestures), we cannot provide all the information at the same time and we are forced to form a sequence: provide the informational elements one by one. Because sequencing and communication co-occur so clearly, we do not want to rule out the influence of communication and communicative pressures in the gesturing experiment.

To sum up, Goldin-Meadow *et al.* (2008) have initiated an interesting experimental method which shows that people bypass the grammatical rules of their native language when they are asked to communicate manually. Goldin-Meadow *et al.* (2008) tested one kind of events: motion events. Langus and Nespor (2010) investigated how more complex events were gestured. In the next section, I will introduce *intensional events*, a class of events that have the same level of complexity as motion events, but that are semantically different from motion events. Below, I will show that intensional events behave differently from motion events in the improvised communication task.

5.3 Motion events vs. intensional events

5.3.1 An intuitive difference

Goldin-Meadow et al speculate that Actor-Patient-Act (ArPA) 'may reflect a natural sequencing for representing events.' As an explanation for this particu-

[5] As noted in chapter 4, advocates of a strong account of Fodor's Language of Thought hypothesis will not agree at this point. I will discuss the role of communication further at the end of this chapter.

lar order they quote results from related research, showing that the Actor and Patient might be situated before the Act, because entities are cognitively more basic and less relational than actions (Gentner and Boroditsky, 2001). Moreover, Patients and Acts are cognitively tied (Goldin-Meadow, 2003), which would link Patient to Act, and result in the ArPA order for gesturing (Goldin-Meadow *et al.*, 2008, p. 9166).

This explanation seems intuitive, especially for the particular kind of events that were used in the experiment described in Goldin-Meadow *et al.* (2008), namely *motion events*. But are motion events representative for all possible events? I will claim that there is a category of events, *intensional events*, for which the ArPA order is not as intuitive as it is for motion events. Intensional events differ from motion events semantically and we will hypothesise that different semantic properties lead to different gesture orderings in the improvised communication task. But first, let us focus on the differences between motion events and intensional events.

The events used in Goldin-Meadow *et al.* (2008) are all events in which someone does something to someone or something else. In these situations, the ontological status of Actor and Patient are similar. E.g., in the example of a girl covering a box, we can summarise the situation as follows: (1) there is a girl, (2) there is a box, and (3) the girl covers the box. In other words, in order for a sentence describing this situation to be true, both the subject and the direct object need to exist, and they need to relate to each other in the right way. Let us compare this situation with a situation in which *a princess wants an apple*. In this example, the ontological status of the Actor and the Patient are not equal: in order for a sentence describing this situation to be true, we need the princess to exist, but the 'ontological demands' on the apple are different: a princess can want an apple without the actual apple being around, or she can want an apple but not one in particular. It is even possible for the princess to want something that does not exist at all, as in 'The princess wants a unicorn.'

The crucial difference between the two events thus resides in a difference between the *verbs* that describe the actions going on in them: both 'cover' and 'want' are transitive verbs (they occur with a subject and a direct object), but 'cover' is an extensional verb, whereas 'want' is an *intensional* verb. Other examples of intensional verbs are 'seek', 'admire', and arguably also 'see' and 'draw'. In the literature, interesting properties of intensional verbs have been described and inventories of intensional verbs have been drawn up (Forbes, 2010; Moltmann, 2008). Let us have a look at them.

5.3.2 Defining intensional transitive verbs

The terms 'intensional' vs 'extensional' are used because of the role that extensions (the object a term refers to) and intensions (the meaning of a term) play in the interpretation of these verbs. In order to interpret an extensional verb, the extension of its complement (the direct object) is important, whereas for the interpretation of intensional verbs, the extension of its complement is less

important than its meaning. This is a first, intuitive definition. But how are intensional verbs defined exactly? In order to get a more precise characterisation of the differences between extensional and intensional transitive verbs, I will give a brief overview of three 'marks' of intensionality that were described in Forbes (2010): substitution-resistance, the availability of unspecific readings, and existence-neutrality.

Substitution-resistance

In sentences with extensional verbs, it is possible to substitute the direct object with one that refers to the same object, without changing the truth value of the sentence. This is illustrated in the following examples:

(1) John lives next to Mark Twain.

(2) John lives next to Samuel Clemens.

Because Samuel Clemens is Mark Twain, sentence 1 is true in exactly the same situations as sentence 2. If we substitute the extensional verb 'live next to' for an intensional verb, 'admire', this is no longer possible.

(3) John admires Mark Twain.

(4) John admires Samuel Clemens.

It might be the case that sentence 3 is true, but that John does not realise that his grumpy neighbour Samuel Clemens is Mark Twain. In that case, 'John admires Mark Twain' is true, while 'John admires Samuel Clemens' is false.

The availability of unspecific readings

In sentences with intensional verbs, it is possible that the direct object remains unspecific. An example of this is the following sentence:

(5) Mary seeks a man.

For this sentence, an interpretation is possible where Mary seeks a man, but not one man in particular. Contrast this with the verb 'kiss'. We cannot say

(6) Mary kissed a man, but not one in particular.

Existence-neutrality

In sentences with intensional verbs, it is possible for the direct object not to exist at all. By contrast, for extensional verbs, the direct object needs to exist in order for the sentence to make sense. It is possible to *seek* a unicorn, but not to *stumble across* one.

5.3.3 Subclasses of intensional verbs

Intensional verbs will always show at least one of the three properties described above. Some verbs manifest all three kinds of behaviour, but there are many verbs that meet only one of the criteria. For certain verbs, it is not always clear whether a particular criterion is met.

The following subclasses of intensional verbs can be distinguished:

- 'Classical' intensional verb: *search for*
- Psych verbs: *dream of, think of*
- Perception verbs: *hear, see*
- Creation verbs: *build, draw, knit, sculpt*

The class of intensional verbs is thus a rather diverse group and there is no general agreement on either the names of the subcategories, the verbs that should be included in them, or even whether all categories listed above are truly intensional.

One might, for example, question the intensional properties of perception verbs: if John *sees* a house, doesn't he just see an existing external object? But the fact that a sentence like 'When John listened to a cello, he heard a violin' is possible, shows that there is, after all, something special about complements of perception verbs. In support of this view, see the following characterisation of perception verbs, as presented in Moltmann (2008):

> The complements of perception verbs [. . .] do not describe the external object that may be perceived, but rather the way the perceived object appears.

In other words, for perception verbs, like for other intensional verbs, it is not the external object that is important for their interpretation, but the intension. This shows that perception verbs have at least an intensional flavor.

A special class of verbs that have been discussed in the literature are creation verbs. To give a very brief summary of this discussion (Zucchi (1999) provides more details), there are two opposing intuitions. The first is reflected in Bennett (1977),[6] who states that when a sentence like 'John is building a house' is uttered, this has a reading in which the house does not exist. The fact that there is this reading, Bennett claims, should be explained by analysing the verb 'build' as intensional.

Parsons (1990), appeals to the following examples, to argue that creation verbs are different from intensional verbs:

(7) Mary built a house.

(8) Mary looked for a unicorn.

[6] Cited in (Zucchi, 1999, p. 190).

Whereas it is possible that the unicorn in sentence (8) does not exist, the house in sentence (7) must exist. Therefore, the class of intensional verbs and that of creation verbs must be different classes. When analysing a progressive sentence like 'Mary is building a house,' he maintains that for this sentence to be true, the house should exist; and similarly so for the circle in a sentence like 'John is drawing a circle,' because "[p]eople do refer to unfinished houses as houses and even—though more reluctantly—to unfinished circles as circles" (Parsons, 1990, p. 178).

It is clear now that especially in the case of creation verbs it is not clear whether they are really intensional in nature. Below, when we will look at the behaviour of intensional verbs in an experimental setting, we will take the widest definition of 'intensional' and include all verbs that have an intensional flavor. A more fine-grained analysis of creation verbs versus other intensional verbs will be given in section 5.4.3.

5.3.4 Intensional events and gesturing order

We have described semantic differences between two kinds of events: extensional (motion) events and intensional events. How will the two behave in the improvised communication setting? An interesting link between semantic properties and word order is provided in Jackendoff (2002). Jackendoff suggests that in simple language systems without full syntax, semantic principles play an organising role in short utterances. We will assume that the gesture strings produced in the improvised communication task are such a 'language system without full syntax', and thus hypothesise that *semantic properties of events are important in the improvised communication setting and will influence the order of the gesturing.*

Let us look again at the event *princess wants apple*. This event is typically described with an intensional verb, and can thus be called an intensional event. As was pointed out above, there is something special about the direct object in such events: the apple in the 'want-event' is not necessarily a concrete object. If we would make a step-by-step analysis of the event, it would look like this: (1) there is a princess, (2) there is something the princess wants, and (3) that is an apple. This analysis reflects the fact that there is something special about the Patient: it is in some sense dependent on the Agent and the Act. The hypothesis we advance is that the semantic differences between extensional and intensional events lead to different orderings when these two classes of events are linearised for communication.

To test this hypothesis, we set up an improvised communication experiment where people are asked to convey the meanings, in an improvised manner, of both motion events and intensional events. We predict that for motion events, which are all typically described by extensional verbs, participants will use ArPA order, similarly to what was shown in Goldin-Meadow *et al.* (2008). In intensional events, however, there is something special about the ontological status of the Patient. We have seen above that the direct objects in such events

are less concrete. Based on the observation in Goldin-Meadow *et al.* (2008, p. 9166) that information that is cognitively more basic is expressed before more complex information, I predict that the Patients in an intensional event will be placed at the end of the gesturing sequence, resulting in a gesturing order of ArAP. In order to make sure that the influence of the native language of the participants is ruled out, we will test participants of two native languages: Turkish (which has SOV as dominant order) and Dutch (which has SVO as dominant order).[7]

5.4 Experiment 1: gesturing motion events and intensional events

This experiment tested the influence of the semantic properties (motion vs intensional events) of an event on gesture ordering in improvised communication.

5.4.1 Method

Participants

16 participants (5 male and 11 female) were recruited from Utrecht University and the Utrecht School of the Arts in Utrecht, the Netherlands. All were native speakers of Dutch (which is an SVO language), and none of the participants had any knowledge of a conventional sign language.

19 participants (10 male and 9 female) were recruited from Bogazici University in Istanbul, Turkey. All were native speakers of Turkish (which is an SOV language), and none of the participants had any knowledge of a conventional sign language.

All participants (Turkish and Dutch) received a small monetary compensation for participating.

Items

The set of items consisted of 20 pictures of motion events (e.g. 'Pirate throws guitar', 'Princess carries vase'), and 20 pictures of intensional events (e.g. 'Cook thinks of sock', 'Leprechaun sees tall building'). Each motion event had a corresponding intensional event, with the same actor and patient, but a different action.

All actors (subjects) in the pictures had particular external characteristics (e.g. a princess with a crown, a pirate with a hat), in order to encourage participants to really gesture all elements in the picture. All patients (direct objects)

[7]Strictly spoken, Dutch is not an SVO language; it belongs to the subclass of V2 languages. The difference between SVO and V2 only surfaces in certain constructions, and not in those used in the experiment. In other words, the events used in the experiment, when described in Dutch, will be described in SVO order and for this study Dutch can safely be used as an SVO language.

Figure 5.1: Two example items: *'pirate throws guitar'* and *'cook thinks of sock'*

were inanimate objects.[8] All pictures had been pre-tested for clarity. Each picture was shown either in its original version or as a mirror image, to control for the left-to-right order of the elements in the pictures.

For the experiment, two versions were created, each consisting of 10 pictures of motion events and 10 pictures of intensional events. The items were presented in random order.

Procedure

Participants were shown pictures of events on a computer screen. They were asked to convey the meaning of each picture to the experimenter (who could not see the computer screen), by using only gestures and no speech. Each picture remained visible on the screen while the participant was gesturing. Participants were told to keep gesturing until they thought they had conveyed the meaning of the picture; no information was given about the amount of gestures to be used.

Before the actual experiment started, participants were shown four practice items. During the practice stage of the experiment, the experimenter gave feedback about whether or not she understood which meaning was conveyed. No spoken feedback was given during the experiment.

After the gesturing part of the experiment, participants were shown the pictures again, and were asked to describe each event using a declarative sentence

[8]This was done to exclude any effects of animacy as described in Meir *et al.* (2010), in which for events with two animate entities, ArAP gesture order was found as the dominant order.

in their native language.

5.4.2 Data analysis and results

The video recordings were coded for gesturing order by two independent coders (80,6% agreement). All gesturing sequences for which there was no consensus were filtered out (62 of 320 recordings). Occasionally, participants produced gesture strings describing an action that did not match with the intended action on the picture; these were removed as well (10 recordings).

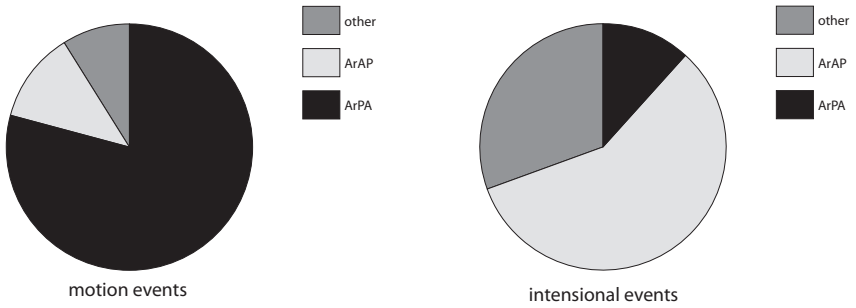


Figure 5.2: Proportions of gesturing orders observed in experiment 1: for motion events, a high proportion of ArPA was observed, and for intensional events, a high proportion of ArAP was observed.

The overall results are shown in figure 5.2. The chart shows the percentages of ArPA strings, ArAP strings and other orders. The category of 'other' orders (which is only a minority of the totality of strings) consisted of strings like PArA, AArP, or strings with either less or more than three gestures. It is clear from the graph that the majority of motion events were gestured in ArPA order, whereas the majority of intensional events were gestured in ArAP order.

The data were analysed using a repeated measures ANOVA. The within subject factors were Picture-type (intensional or motion) and Order (ArPA or ArAP); the between subjects factors were Version (version 1 or 2) and Language (Turkish and Dutch). We found a significant interaction between Picture type and Order: $F(1, 31) = 268.911$, $p = .000$. No significant interaction was found between the main interaction and the effect of version ($p = .836$) or language ($p = .208$).

Pairwise Bonferroni corrected comparisons reveal that among gesture strings of motion events, the proportion of ArPA order was high ($M = .714$, $SE = .036$), whereas the proportion of ArAP order was low ($M = .106$, $SE = .035$).

Among gesture strings of intensional events, the proportion of ArAP order was high ($M = .576$, $SE = .046$), whereas the proportion of ArPA order was low ($M = .119$, $SE = .024$). See figure 5.3.

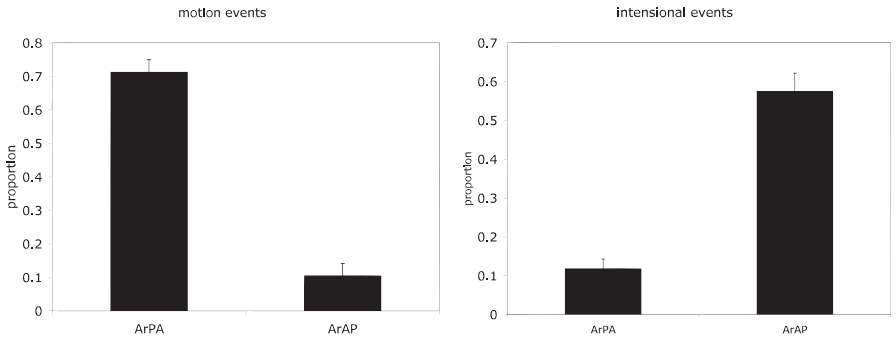


Figure 5.3: Mean proportions of ArPA and ArAP gesturing orders for motion events (in the first graph) and intensional events (in the second graph). Error bars indicate standard error of the mean.

5.4.3 Post hoc analysis: creation verbs

Above we have seen that in the literature, there is quite some discussion on the question whether creation verbs like *draw, build* and *sculpt* are truly intensional in nature. Because we cannot exclude the possibility that creation verbs are *not* intensional, it is interesting to compare the gesturing sequences for creation verbs with those of other intensional verbs.

We conducted a repeated measures ANOVA, using the within subject factor Picture-type (creation events and other intensional events) and between subjects factors Version (version 1 or 2) and Language (Turkish or Dutch). We found that Picture-type significantly influenced the proportion of ArAP ordered strings: $F(1, 29) = 4.679$ and $p = .039$. Pairwise Bonferroni corrected comparisons show that the proportion of strings in ArAP order was smaller when people were communicating about a creation event ($M = .445$, $SE = .064$) than when they were communicating about other intensional events ($M = .602$, $SE = .058$). There was no significant interaction with the main effect of Version ($p = .787$) or Language ($p = .138$).

Thus, the proportion of ArAP gesturing strings for creation events is significantly lower than of those for other intensional events. However, ArAP is still the most commonly used string for creation verbs, just like for other intensional verbs. In other words, creation events behave like other intensional events, but less consistently so.

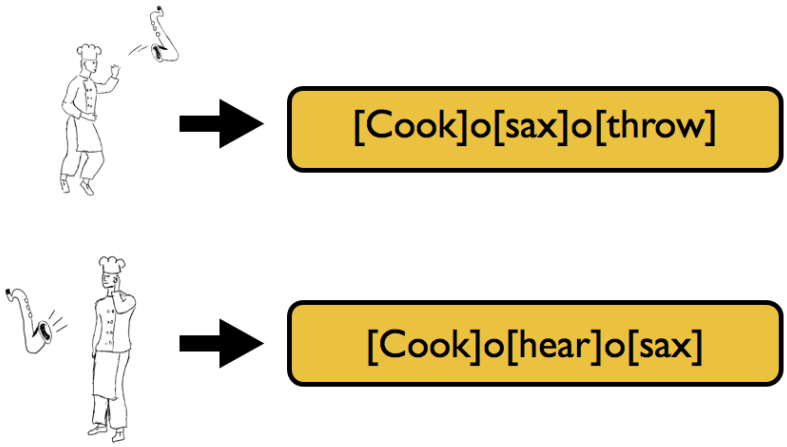


Figure 5.4: Different actions giving rise to different gesturing orders: ArAP for intensional events and ArPA for extensional events.

5.4.4 Discussion

For motion events, the results obtained in Goldin-Meadow *et al.* (2008) were replicated, but for intensional events, different gesture sequencing was observed: the Patient was placed *after* the Act, resulting in an ArAP order (see figure 5.4). This confirms our hypothesis that semantic properties of the verb (or more precisely, the action depicted) influence the ordering of elements in improvised communication.

5.5 A word on methodology

Using gesture in a lab setting in order to make people communicate in an improvised manner is a relatively new approach. Hence, it is interesting to focus specifically on the methodology used in these experiments. That is what I will do in this section: I will review some of the methodological problems in Goldin-Meadow *et al.* (2008) and Langus and Nespor (2010). For each, I point out how we have adapted the setup in our experiment to overcome these problems and improve the experimental design.

5.5.1 Methodological issues: Goldin-Meadow *et al.* (2008)

As described above in section 5.2.1, vignettes depicting motion events were used in the experiment described in Goldin-Meadow *et al.* (2008). Of the forty vignettes, 20 were intransitive (i.e., ones that are typically described by a sentence

with an intransitive verb), and 16 were 'transitive actions'. Of the transitive actions, 8 involved movement in place (e.g., 'boy tilts glass', or 'captain swings pail'), and 8 involved movement crossing space (e.g., 'man moves garbage can to motorcycle man', 'girl gives flower to man'). The items depicting movement across space depict, apart from an agent, a patient and an act, also an *endpoint* (described by the indirect object), e.g., *man* in 'girl gives flower to man'. These items thus contain four elements instead of three, which makes the gesturing of these events more complicated, and that is reflected in the results: relatively few participants included the Patient in their gesturing sequence for this class of events. We therefore decided not to use any vignettes depicting movement across space.

Omission in the gesture sequence of one of the three elements in the picture was a problem that occurred not only for the events crossing space, but also for other events. On page 9167 in Goldin-Meadow *et al.* (2008), it is noted that of the results, 501 gesture strings contained two relevant elements, and only 113 contained three.

One of the reasons for this might be that in the setup of the experiment, both the experimenter and the participant were able to see the screen with a picture of the event. Due to this setup, there was no strict necessity for the participants to really communicate all the details of the event depicted. In our experiment, the placement of the screen was such that only the participant could see the screen.

In some of the gesture strings, the patient and the action were gestured in one instead of two separate gestures, which is another cause of the high percentage of two gesture strings. A closer look at the items used in the experiment reveals that some of the items trigger this strategy more than others. Examples of items for which patient and act were collapsed into one gesture are 'boy stirs spoon' and 'man plays guitar' (Goldin-Meadow *et al.*, 2008, table S1). When people gesture a guitar, they usually do this by 'playing' it (and not by, e.g., indicating its shape). Thus, the action (playing) and the patient (guitar) are taken together in one gesture. In our experiment, we solve this issue by taking combinations of action and patient that are not prototypical combinations. For example, instead of using 'man plays guitar' as an item, we use 'man throws guitar' (see figure 5.1).

A third reason for the omission of elements in the gesturing might be that the subject is easily forgotten. We will say more about this below, when discussing methodological issues in Langus and Nespor (2010).

5.5.2 Methodological issues: Langus and Nespor (2010)

In the construction of their experiment, Langus and Nespor (2010) have taken into account some of the problems that occur in Goldin-Meadow *et al.* (2008). In their first experiment, Langus and Nespor use 32 drawn vignettes depicting events in which 'someone does something to someone or something else' (Langus and Nespor, 2010, p. 5). In other words, all of their items depict

transitive actions. They remark that in each vignette, 'each of the three constituents unambiguously matched the category of the Subject, the Object or the Verb.' (Langus and Nespor, 2010, p. 5). This means that for an event like 'a girl catches a fish', the opposite, 'a fish catches a girl' is not a viable option. Moreover, all constituents of the events (subject, object and verb) were equally frequent, and occurred in different combinations with other constituents during the experiment. We decided to use the same approach in our experiment.

We do find some potential problems in their experimental setup, however. First of all, participants in the experiment were instructed to describe each vignette using 3 gestures. We think that this instruction is too informative and might prompt participants to think about the purpose of the experiment.

Despite the explicit instruction to use three gestures, many participants in Langus and Nespor's experiment used 2-gesture strings: the gesture strings contained all three elements in only 58.6% of the cases in Italian speaking participants and 63.2% in Turkish speaking participants (Langus and Nespor, 2010, p. 295). In two-gesture strings, the Subject was omitted most often. In our experiment we avoid the issue of omitted Subjects, by giving all Subjects specific characteristics, like a pirate with an eye patch and a witch with a big hat.

5.5.3 Representing events in images

In Goldin-Meadow *et al.* (2008), short animations were used to represent the events that were to be gestured by the participants. In Langus and Nespor (2010), on the other hand, only pictures were used to represent the events. In our experiment we use pictures only, for practical reasons. It is very well possible to represent events by means of a picture, but if we want the participants to understand which event is meant, it is worthwhile to think carefully about how a picture best represents an event.

There are many ways to analyse events and to divide the class of events in subclasses. One of the ways, put forward in Vendler (1957), is to distinguish between four types of eventualities: states, activities, accomplishments, and achievements.

> An activity, such as John's walking uphill, is a homogeneous event: its sub-events satisfy the same description as the activity itself and it has no natural finishing point or culmination. An accomplishment, such as John's climbing the mountain, may have a culmination, but is never homogeneous. An achievement, such as John's reaching the top, is a culminating event (and is therefore always instantaneous). And a state, such as John's knowing the shortest way, is homogeneous and may extend over time, but it makes no sense to ask how long it took or whether it culminated. (Casati and Varzi, 2010)

Most of the items used in the experiments by Goldin-Meadow *et al.* (2008) and Langus and Nespor (2010) fall into the category of either activity or ac-

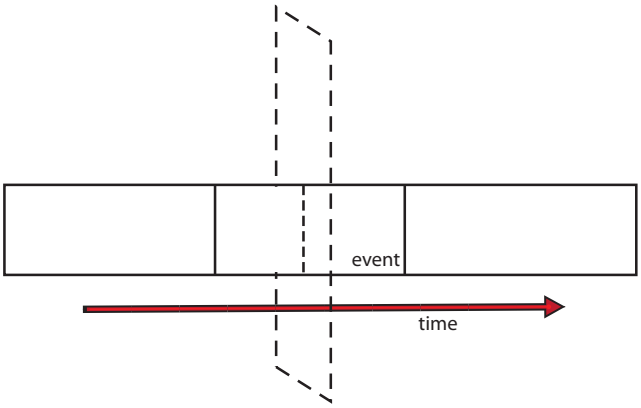


Figure 5.5: A snapshot of an event

complishment. Looking at the pictures used by Langus and Nespor, we must realise that in a picture, only a snapshot of the event is given (see figure 5.5). What snapshot exactly is taken to represent the whole event might have an influence on how the event is going to be described by the participant.

For activities, taking a snapshot should not pose any problems. For example, pushing a cart is a homogeneous process. Whichever snapshot we make of this activity, we will see a representative picture of what is going on.

Taking a snapshot of accomplishments might result in problems. Some accomplishments have a resultative state. For example, catching a fish has as a resultative state the possession of a fish, or throwing a ball into a box has as a result that the ball is now located inside the box. If we want people to act out such events, we do not want them to focus too much on the resultative state (for the simple reason that we would like to see what they do with the actor, patient and act in the event itself, not in the resultative state).

Some accomplishments, on the other hand, have a preporatory phase. For example, shooting an elephant involves aiming for the elephant, and aiming is something quite different from shooting. So it is important, when depicting an accomplishment with a single image, to choose a snapshot that is still within the action itself, to avoid too much attention being paid to either the result of the action or the preparation. The figures we used in the experiment are listed in Appendix 7.3. An example is the event in which a princess cuts a scarf. The result state of this event would be a scarf in two pieces. To make sure that participants will not start to attempt to gesture that, we depicted the event at a stage where there is still clearly one whole scarf.

Of course, this is not a recipe for making a perfectly representative image of every kind of event, but keeping the properties of processes and accomplish-

ments in mind when preparing images for the experiment certainly helps to produce effective images.

5.5.4 From production to interpretation

To wrap up, we have seen that events with different semantic properties (extensional vs. intensional) give rise to different gesturing orders (ArPA vs. ArAP). Interpreting this along the lines of the theory presented in chapter 4, this would mean that people linearise mental representations of events differently in different circumstances: when a Patient is more abstract, or dependent on the Act, the information is linearised in an ArAP order, and an ArPA order is chosen otherwise. I have claimed that linearisation of information, going from a non-linear mental representation to a string, is something that is inherently communicative. In other words, the order that is imposed on utterances is done so *for communicative reasons*: to make it easier for hearers (observers) to get the message. If this is indeed the case, we would expect to see an influence of order on the interpretation of gesturing sequences. This hypothesis is tested in the next section.

5.6 Experiment 2: the interpretation of gesture strings

It has become clear from the experiment described above that the semantic differences between motion events and intensional events have an influence on gesture production in the improvised communication task, resulting in ArPA order for extensional (or motion) events, and ArAP order for intensional events. I have claimed that these orders are the result of a process of linearisation, in which a non-linear mental representation is forced into a string, which has linear order, and that this process is a communicative process (see chapter 4). If that is true, and the formation of utterances in the improvised communication task is indeed a communicative process, then we would expect to see an influence of the ordering of elements in a string on the interpretation of this string.

In order to test this, we set up an experiment in which participants were asked to watch videos of gesture strings. Each video contained a string of gestures with three elements: an Actor, a Patient and an Act. The essential element in this experiment is that the recorded Acts were ambiguous. We recorded gestures that could be interpreted in two ways: as an Act creating an extensional event, or as an Act creating an intensional event. An example is depicted in figure 5.6: this Act can be interpreted as *build* or as *climb*, because the gesturing is somewhat vague.

Using the ambiguous Acts, we created a series of ambiguous gesturing sequences in two orders: ArPA and ArAP. In the production experiment, ArPA order was used for motion events, and ArAP was used for intensional events.

Under the assumption that our hypothesis that this was done for communicative purposes is true, we expect ArPA strings to be interpreted as motion events and ArAP strings as intensional events (van Leeuwen, 2010).

5.6.1 Method

Items

We created short movie clips showing an actor gesturing simple events. The verbs were gestured in such a way that the events acted out could be interpreted either as a motion event or as an intensional event.

An example of the way in which these ambiguous items were gestured is given in figure 5.6, in which the ambiguous Act 'climb/build' is shown. Thus, the gesture in the figure can be interpreted as a climbing action, as well as a building action. We created two videos of each ambiguous verb by adding a gestured subject and object to the transitive event. There were two different orders: [Actor-AmbiguousAct-Patient] and [Actor-Patient-AmbiguousAct]. Examples of two videos using the same material are shown in figure 5.7 and 5.8. Each video consists of *exactly the same* video material, but the elements are put in two different orders. Because participants were presented the two answering options before watching the video, the Actor and Patient in each video were relatively easy to recognise.

The following ambiguous events were used; for every item in the list, the option marked with m creates a motion event, and the option marked with i creates an intensional event.

- Pirate drops$_m$/searches$_i$ ball.
- Princess breaks$_m$/sculpts$_i$ vase.
- Leprechaun cuts$_m$/draws$_i$ pizza.
- Witch eats$_m$/wants$_i$ banana.
- Witch paints$_m$/paints$_i$ table.[9]
- Girl sleeps on$_m$/dreams of$_i$ book.
- Girl kisses$_m$/thinks of$_i$ doll.
- Princess talks to$_m$/talks about$_i$ snowman.
- Pirate throws$_m$/hears$_i$ guitar.
- Cook stirs$_m$/smells$_i$ soup.
- Leprechaun hits$_m$/feels$_i$ book.
- Witch climbs$_m$/builds$_i$ house.

Of the 12 video pairs, two versions (Version 1 and 2) were created. The videos were randomly categorised as belonging to version 1 or 2, taking care that each version consisted of 6 videos in ArAP order, 6 videos in ArPA order. Four

[9] In the first interpretation a witch painting an existing table is meant; in the second a witch painting a table on a canvas.

Figure 5.6: An ambiguous action: 'climb' or 'build'.

Figure 5.7: An ambiguous gesture string in ArAP order: witch-climb/build-house. Each still represents an element: 'witch' was gestured by indicating the shape of the hat; 'house' was gestured by indicating the shape of a rooftop.

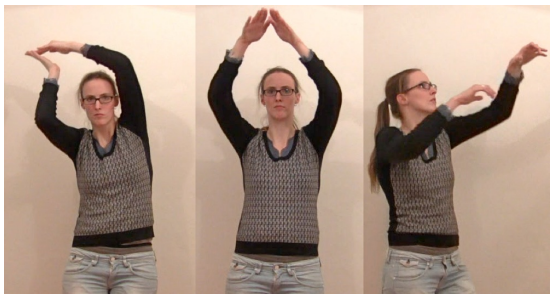


Figure 5.8: An ambiguous gesture string in ArPA order: witch-house-climb/build. The video material used for this video was exactly the same as for the video depicted in 5.7; only the order is different.

fillers, items with unambiguous actions, were added to each version. The videos were shown to participants in a two alternative forced choice task; pictures of the corresponding intensional and extensional (motion) events were shown as the two answer possibilities.

Participants

Forty one native speakers of Dutch (16 male, 25 female) were recruited from the Utrecht University library in Utrecht, the Netherlands.

Forty native speakers of Turkish (12 male, 28 female) were recruited from the library at Bogazici University in Istanbul, Turkey.

The participants did not receive a monetary compensation for participating.

Procedure

The participants were shown videos on a laptop screen and were asked to choose, after each video, the picture that fitted best with the event acted out on the video. First two practice items with unambiguous verbs were shown, followed by the ambiguous items and fillers. The ambiguous items and fillers were presented in random order. The two answer possibilities were shown before each video,[10] and again afterwards. The order of the two answer possibilities was also randomly determined.

The experiment was conducted on a laptop in the library, and took about ten minutes to complete.

5.6.2 Data analysis and results

Upon re-analysis of the video clips we decided to exclude two videos from the results: 'Pirate drops$_m$/searches$_i$ ball' and 'Girl kisses$_m$/thinks of$_i$ doll'. These two videos differ from the others in the sense that the ambiguous actions they depict consist of two sub-gestures,[11] whereas for all other ambiguous actions, only one gesture is used.[12]

The data was analysed using a repeated measures ANOVA. The within subjects factor was Order (ArPA or ArAP gesturing order), and the between subjects factors were Version (version 1 or 2) and Language (Turkish or Dutch). We found a significant main effect of Order: $F(1, 77) = 23,454$, $p = .000$. No significant interaction was found of Language ($p = .691$). However, a significant interaction on the main effect of Version was found ($p = .004$). The latter might be an (undesired) result of the fact that some videos showed a bigger effect than

[10] A pilot had pointed out that without showing the two answer possibilities before the video, the task was too hard (participants were unprepared when watching the video, and were not able to make a choice after the video).

[11] A drop-gesture followed by a search gesture for the former, and a think of gesture followed by a kiss gesture for the latter.

[12] Including the two deleted item in the analysis still yields significant main effect of video order: $F(1, 77) = 17.798$, $p = .000$.

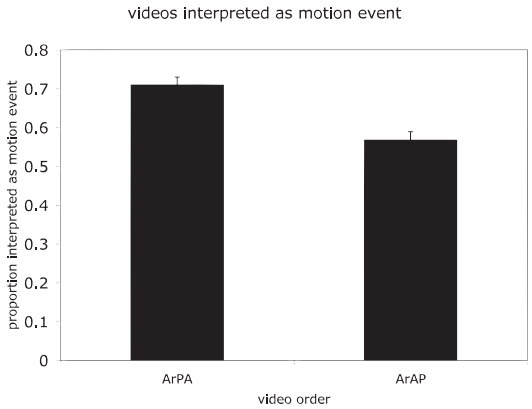


Figure 5.9: Mean proportions of videos interpreted as motion event for ArPA and ArAP video order. Error bars indicate standard error of the mean.

others. Given the fact that the two versions were created randomly (see above on page 144), this effect is harmless.

Pairwise Bonferroni corrected comparisons show that for ArPA gesturing sequences, a motion answer was chosen more often ($M = .711$, $SE = .019$) than for ArAP gesturing sequences ($M = .569$, $SE = .020$). See figure 5.9.

5.6.3 Discussion

The two different gesturing orderings led people to interpret the gesturing differently: ArPA gesturing strings were more likely to be interpreted as motion events than SVO strings were, and ArAP gesturing strings were more likely to be interpreted as intensional events than ArPA strings were. So we are safe to conclude that the order of gesturing has an influence on the interpretation of the gesture strings. This supports our hypothesis that choosing a different ordering in order to convey different kinds of meanings serves a communicative purpose.

The effect of order in this experiment was not as strong as the effect found in the production experiment. One of the reasons for this might be that this experiment was conducted, not in a lab, but in a public place, the university library. This may have led to more noise in the data.

Moreover, when choosing an interpretation for each video, participants could do this on the basis of many factors, because they were not instructed to make their choice on the basis of the gesturing order. In other words, participants may have based their choice for a particular interpretation of a video, not on the gesturing order, but for example on the shape of the gestures or the facial expression of the actor. This may have added noise to the data as well.

Despite these noise factors a significant difference was found, which makes the results even more striking.

5.7 General discussion

5.7.1 Overview of the results

Section 5.4 showed that the semantic properties of events have an influence on the gesturing order in improvised communication: ArPA order is used for extensional (motion) events, and ArAP order is used for intensional events.

Section 5.6 showed that when people interpret gesturing sequences, the ordering of the gestures influences the way they interpret these events: for gesturing sequences with ambiguous actions, an extensional interpretation was chosen significantly more often after an ArPA sequence than after an ArAP sequence.

5.7.2 Word order in communication

In previous publications, similarities were pointed out between the improvised communication task and newly emerged sign languages like Al Sayyid Bedouin Sign Language and Nicaraguan Sign Language (Goldin-Meadow *et al.*, 2008, p. 9167). These similarities led to the hypothesis that results from the improvised communication experiment can tell us something about the emergence of language in general: it was argued that ArPA word order "may reflect a natural disposition that humans exploit not only when asked to represent events nonverbally, but also when creating language anew" (Goldin-Meadow *et al.*, 2008, p. 9167). This in turn connects well to the claim made in Newmeyer (2000), that the earliest human language had rigid SOV word order.

The production experiment described in this paper shows that it is not ArPA word order as such that is important in the improvised communication task, but rather the meanings that are to be conveyed: extensional events lead to ArPA ordering, whereas intensional events lead to ArAP ordering. This shows that in emerging communication systems, meaning and structure have more to do with each other than previously thought. Moreover, it suggests that ordering information in utterances in these systems is quite an active process, rather than simply a reproduction of how information is represented mentally. To repeat my analysis from chapter 4, what essentially happens in the experiment is the following: people are presented with a 2-dimensional image of an event, and then they are forced to linearise the information depicted in the image into a sequence. When making a sequence of the different elements, they are forced to impose an order on the information. So it is only in making the information *public*, in being involved in *communication*, that ordering plays a role. As is shown in the experiment in section 5.4, people show systematicity

in the ordering they choose. This systematicity can be seen as the beginnings of, or precursor of, syntactic rules.

The experimental results suggest that semantic properties of the events communicated about in improvised communication play a role in the structuring of the information, and that this is essentially a communicative phenomenon. But communication is a process that has two perspectives: that of the speaker, and that of the listener, or interpreter. If the suggestion formulated here is right, then the relation between two kinds of events and the two different orderings should play a role, not only on the speaker side of the communicative process, but also on the side of the interpreter.

The second experiment described in this chapter shows that the order of gesturing does indeed have an influence on the interpretation of the gesture strings. This supports the hypothesis that the distinction found in gesture production between ArAP and ArPA ordering for intensional events and motion events respectively, has a communicative function.

5.7.3 Comparison with restricted linguistic systems

Effects of meaning on structure in simple language systems have been found in other linguistic phenomena where improvisation is required. In chapter 2 we have seen a description of the Basic Variety: in the process of acquiring a second language outside the classroom, adult learners go through a stage that has been characterized as being (1) determined by a small number of organisational principles, (2) largely independent of the source or target language of the learner and (3) simple but relatively successful for communication. Some examples of organisational principles of the Basic Variety are FocusLast ('put the information that is in focus, new information, at the end of the sentence') and AgentFirst ('the NP referent with the highest control comes first'). Similar organisational principles were described for, e.g., pidgin languages (Jackendoff, 2002). We have also seen—in chapter 2—that the fact that the organisational principles described above are found consistently in linguistic phenomena like the Basic Variety and pidgins, and seem to be independent of the native language of their users, makes these phenomena interesting for the language evolution debate: they might tell us something about the structure of evolutionarily early language (Jackendoff, 2002; Schouwstra, 2010).

As described in chapter 4, the circumstances under which people create systems like the Basic Variety and pidgins are very particular and also often undesirable. Moreover, the circumstances under which these restricted systems emerge are not controllable, and the data is therefore not very clean. Finally, the focus of data collection for these systems was mainly on production, and not on comprehension. Therefore, it would be valuable if we would be able to collect this kind of data in a controlled environment, like in a laboratory. The improvised communication task provides us with exactly this: if we see the task as a setting where restricted linguistic systems are produced, we can obtain data in the ideal way described above. The improvised communication

task can be seen as an environment where restricted linguistic systems are produced, because the setting shares many properties with the settings of existing restricted linguistic systems. In the experiment, like in those situations, subjects cannot use their native language to express themselves and are forced to improvise, using whatever they have in their restricted inventory. The improvised communication task thus offers us a way to collect data about the earliest stages of language emergence in a controlled manner.

Moreover, the view of the improvised communication task as a restricted linguistic system allows us to compare the influence of semantic structures in existing studies into, e.g., the Basic Variety and pidgin languages to that in lab situations. Thereby, one can go back and forth between data from existing linguistic systems and lab data from the improvised communication experiment, using both phenomena to formulate sharper hypotheses about the mechanisms governing protolanguage. In the previous chapter, we have made observations about existing restricted systems concerning dominant word order. Let us now compare these results to the lab results obtained in this chapter.

In section 4.4.2, it was observed that in existing restricted linguistic systems, two word orders that seemed dominant are SVO and SOV. These correspond with the ArPA and ArAP orders observed in this chapter. However, the effect observed in the experiments described in this chapter of a distinction between intensional and extensional events, has not been observed in existing restricted linguistic systems. This might have several reasons. For pidgin and Basic Variety, we have seen that the word order of simple expressions may be influenced by the full languages spoken by the speakers in these situations (Substrate/superstrate language in the case of pidgin, and source language in the case of Basic Variety). For newly emerging sign languages, it may even be the case that there is some effect of the intensional/extensional distinction on word order, but one that has not been observed previously, because no one focused specifically on this effect. Finally, a crucial difference between the lab experiment and the existing restricted systems is that participants in the lab study have no existing conventions *whatsoever* to rely on: they are presented with a task they have never done before. Speakers of pidgin, or Homesign, for example, do have the beginnings of an established system of conventions to rely on (see chapter 4, for a characterisation of pidgin in terms of emerging conventions). It might very well be the case that these emerging conventions prevent the usage of different word orders for different situations. We will take this issue up in chapter 7, where future research concerning this issue is sketched.

Furthermore, it was observed that in none of the newly emerging sign languages (homesign, ABSL and NSL) was SVO found as a dominant word order. It was suggested that this might have to do with the modality: it could be that in the manual modality it is simply only possible to first describe the participants in an Act, before describing the Act itself. The first experiment in this chapter has shown that this is not the case: when describing intensional events, participants first describe the Act and then the Patient.

Finally, the principle AgentFirst, that was suggested, in Jackendoff (2002), as a governing principle for the Basic Variety and for pidgin, is very clearly also at work in the lab experiment: almost without any exception, all gesturing strings started with Actor rather than Patient or Act.

5.7.4 On modularity

In section 5.2.2 we have seen the improvised communication experiments as carried out in Langus and Nespor (2010). They draw the conclusion that:

> [T]he prominence of the SOV and the SVO orders among the world's languages originates from different cognitive systems: SOV is the preferred constituent order in the direct interaction between the sensory-motor and the conceptual system; the SVO order is preferred by the computational system of grammar. (Langus and Nespor, 2010, p. 307)

With the results of our experiments, we offer a different explanation. There is no fixed word order for improvised communication, but in improvisation situations, strings are generated dynamically, and word order has a semantic origin. This explanation connects well to other sources of evidence about the origin of word order, as we have seen in restricted linguistic systems.

Moreover, our experiments offer a better explanation for the results of the second experiment in Langus and Nespor (2010). Let me briefly recapitulate what was shown in this experiment.

In the experiment, the improvised communication of complex events was investigated. Vignettes were used that represent complex situations, such as, e.g., *[the man tells the child [that the girl catches a fish]]*. It was found that both Turkish and Italian participants gesture the subordinate clause after the main clause, an order typical of SVO languages. In Langus and Nespor (2010) it is concluded that the SOV order in improvised situations does not generalise to more complex SOV-language-like constructions and that participants thus did *not* use their computational system of grammar to produce the strings. This interpretation is puzzling, because later in the article they interpret a preference for SVO order as originating from the computational module. Thus, it is not clear in the context of this article, which module is responsible for what and on the basis of which assumptions this is the case.

In my view, the complex situations that were used (all of which involved *thinking* or *telling*) are all instances of *intensional events* and thus it is not surprising that a gesture ordering was found for these complex situations that is consistent with SVO.

5.8 Conclusion

In this chapter, we have shown that when people are forced to communicate about simple events in an improvised manner, this does not necessarily lead to ArPA ordering of utterances. Rather, the semantic properties of the Act that is described are decisive in the ordering of gestures in the improvised communication task.

In previous research (Goldin-Meadow *et al.*, 2008) it was pointed out that it is striking that ArPA order was observed in these non-verbal communication experiments, because this corresponds with SOV word order, and this is one of the two most common word orders in the languages of the world. The order that we have found besides ArPA, namely ArAP, corresponds with the other most frequent dominant word order in the languages of the world: SVO. Thus, it might be the case, but this is pure speculation, that the experiment points out the source of the two most common dominant word orders in the world. But it was already suggested in chapter 4 that it may not be dominant order, but rather *variation* in word order that is an interesting phenomenon in language systems where there is no fully established system of conventions: in which cases do people change the order in which they present information?

The experiments in this chapter support the view that in language systems without full syntax, semantic properties play an organising role, a process that is also seen in other linguistic situations where improvisation is required. Moreover, we have shown that different ordering of the constituents in improvised communication sequences results in different interpretations. This supports the view that choosing different utterance structures in order to express different kinds of meanings has a communicative function.

CHAPTER 6

The temporal displacement strategy

6.1 Introduction: adding temporal information to a proposition

The previous chapter focused on the ways in which semantic characteristics of events play a role in the linearisation of information. In other words, in which order are an act and its participants described in communication? We have shown, in the previous chapter, that thematic relations play a role in the linear organisation of improvised utterances: there is a general tendency to express information about the individual with the role of Agent first; this is consistent with what has been found in many restricted linguistic systems. Moreover, the experiments carried out in the previous chapter showed that the semantic nature of the event has an influence on the linear organisation of the elements in communication: for extensional events Patients are described before the Act, and for intensional events, the Act is described before the Patient (I have abbreviated these orders as ArPA and ArAP). It was shown that this not only works for production in improvised communication, but also for interpretation. An interpreter of improvised utterances uses information from the order of presentation to determine the meaning intended by the speaker. Thus, the way in which the verb and its arguments are ordered in improvised communication has a communicative function.

Thus, chapter 5 investigated the linearisation of propositional information. This chapter investigates what happens when information is *added* to a proposition. I will do this by looking at what happens when one wants to describe, of an event, *when it takes place*. In other words, what happens when I would

like to convey the information that *Fred hit the tree*, but I would like to add the information that this happened *yesterday*?

In chapter 4 we have seen that in full language, one can introduce a temporal frame by putting a temporal adverb in front of a sentence. The temporal adverb specifies the place on the time axis about which something is going to be said. This is in Reichenbach's terms the reference point R. The verb form says something about the relation between R and E, the time of the event. An example to illustrate this is the contrast between *Yesterday I met John*, for which R and E coincide, and *Yesterday, I had met John*, for which E comes before R (in other words, the sentence specifies a point on the time axis after the 'meeting-John-event').

In restricted linguistic systems, there is no verb inflection, so the relation between E and R cannot be expressed. Still, speakers in such systems do express information about the past and the future, and they do this by applying the temporal displacement strategy. The temporal displacement strategy tells a speaker to place a temporal adverb in the initial position of an utterance, in order to specify the reference point R. By lack of verbal inflection, the relation between E and R remains underspecified, and it is therefore assumed that E, the event, takes place *at* R. This means that an utterance like *Yesterday John tree hit* describes the event in which John hit a tree, which took place yesterday.

This chapter investigates, first of all, whether this pattern can be reproduced in a lab experiment. Moreover, we would like to know if the usage of the temporal displacement strategy really is a result of the participants bypassing the rules of their native language. In the previous chapter, we have seen that when people communicate simple propositions in an improvised manner, they by-pass the grammar of their native language. When the expressions become slightly more complex, will people still bypass their native grammar? There are two possible answers to this question:

- Langus and Nespor (2010) suggest that, when engaged in improvised communication, people bypass the grammar rules for their native language when they communicate about simple situations, but that this changes when they communicate about more complex situations: then, they claim, the computational module in a speaker's cognitive system is activated, and they start using grammar rules. See section 5.2.2 for more details about the modular conception of cognition that underlies this view. It is possible that adding temporal information to a simple proposition triggers participants to no longer bypass the grammar rules of their native language.

- The data from existing restricted linguistic systems suggest that when people have to communicate in an improvised manner, they organise their utterances according to semantic and pragmatic principles, even when their utterances become more complex. E.g., de Swart (2009) shows that speakers of Basic Variety bypass the grammatical rules of both their

native language and the target language when they express negative information, and Benazzo (2009) describes various source/target language-independent ways to express temporal information observed in the Basic Variety (one of which is the temporal displacement strategy).

In this chapter, I will show evidence that supports the second of these options. Thus, the behaviour of participants in a lab study on improvised communication replicates the observations from restricted linguistic systems 'in the wild'. This finding strengthens the hypothesis that we can use the improvised communication experiment to collect restricted linguistic data. Moreover, the experiment allows us to look in more detail at the principles that play a role when people construct utterances in an improvised manner. We will explore what happens when temporal information is added to a simple proposition: how are the two kinds of information put together?

In section 6.2 a pilot study is presented and section 6.3 describes a full experiment based on the pilot. Section 6.5 interprets these results, proposes a model of the way in which different levels of information are put together in improvised communication, and sketches the evolutionary implications of this model.

6.2 The displacement strategy in the lab: a pilot study

I set up a pilot study in which people were asked to communicate in an improvised manner, by using only gesture and no speech (similar to experiment 1 in the previous chapter). In the study, the aim was to make people communicate about events that took place in the past or that will take place in the future, and to see if people follow the *displacement strategy* that was observed in other restricted linguistic systems.

In the improvised communication experiment, like in restricted linguistic systems, people are not able to use verbal inflection. It is possible, however, to refer to the past and the future by using temporal adverbs. In English, it is possible to insert a temporal adverb, either at the beginning or at the end of a sentence, like in the following examples:

(1) Tomorrow, the king will take a shower.

(2) The king will take a shower tomorrow.

In the experiment, English sentences were offered, and people were asked to convey the information by using only gesture and no speech. If participants apply the displacement strategy, they will be likely to put the temporal adverb at the beginning of their gesturing sequence.

6.2.1 Method

Twelve participants were recruited from the University of Edinburgh. All were native speakers of English (an SVO language) and none of the participants were familiar with any conventional sign language. The participants did not receive a monetary compensation for participating.

The set of items consisted of 15 spoken sentences. Each spoken item was followed by a three second silence and subsequently a short tone. Of the 15 items, 4 contained temporal information, all expressed by a temporal adverb as well as verbal inflection. The temporal adverbs that were used were *tomorrow, yesterday, at six o'clock* and *at night*. Temporal adverbs were placed either at the beginning ($T_{initial}$), or at the end of the sentence (T_{final}), such as in the example sentences (1) and (2). All sentences in which a temporal adverb occurred were intransitive sentences (although one could argue that 'to take a shower' is transitive, but 'take a shower' was generally gestured with only one gesture).

Each participant received two items under the $T_{initial}$ condition, and two items under the T_{final} condition. The temporal and 'non-temporal' items were presented in random order.

Participants were asked to listen to each spoken sentence through headphones. They were told that each sentence described a simple situation, and that they had to describe each situation by using only gesture and no speech.

6.2.2 Results

Gesturing sequences were observed and written down by the experimenter. No video recordings were made of this pilot study. Of 48 strings with temporal information, 3 were not usable because the participant indicated that he/she did not know how to gesture the temporal information.

Of the remaining 45 temporal gesturing strings, 22 were generated under the $T_{initial}$ condition. Of these 22 strings, 21 had the temporal adverb in the initial position. In other words, participants showed a tendency to put the temporal information in the same position as it was in the recorded sentence. Of the remaining 23 strings under the T_{final} condition, however, only 13 followed the sentence order; in 10 gesture strings, the temporal adverb was placed in the initial position. See table 6.1.

Thus, in improvised communication, there seems to be a tendency, when a temporal adverb is presented in sentence-final position, to deviate from this order and put the temporal adverb in the initial position, thereby applying the temporal displacement strategy.

6.2.3 Discussion

This pilot study was very small, but it already indicates that the displacement strategy may be found in a lab situation.

input: spoken sentences	output: gesture strings	
$T_{initial}$: 22	$T_{initial}$	21
	T_{final}	1
T_{final}: 23	$T_{initial}$	10
	T_{final}	13

Table 6.1: Gesturing orders and their frequencies used by the participants, sorted by input type.

There are, however, some methodological problems, and an issue concerning the interpretation of the results. The latter is the problem that we cannot be sure whether people prefer to put temporal information at the beginning of utterances *only when they are in an improvisation situation*, or that the tendency to front temporal adverbs is a more general tendency. In other words, it could be that the participants would show the same behavior when just using their native language, and we cannot be sure that the behaviour we observe is specific for improvised communication. It would be better if we could ask participants to produce both spoken sentences and gesture sequences, similarly to what was done in Goldin-Meadow *et al.* (2008).

When executing the pilot study, it became clear that not every participant felt comfortable communicating by using only gesture. For those who were most uncomfortable, gesturing about temporal information was especially hard. Some participants in the pilot study indicated that they could not think of a way to convey the temporal information. Of the temporal expressions that were used in the pilot study, the expression *at six o'clock* proved to be the easiest to convey by using gesture. I thus decided that the next experiment should have temporal expressions like this.

Further, the fact that the stimuli were spoken sentences was not ideal. In some cases, it was clear that participants were mentally repeating the sentence to themselves. This could have affected the order of their gesturing. It would thus be better to use pictures instead of spoken sentences, similarly to what was done in the experiment in the previous chapter.

6.3 Experiment: gesturing temporal displacement

6.3.1 Introduction

After observing the problems with the pilot study, some adjustments were made to the design, and an experiment was conducted with more items and more participants. We chose to use pictures—instead of spoken sentences—as stimuli, so that we could ask the participants to first describe the items using speech and then using only gesture (similarly to what was done in Goldin-

Figure 6.1: Example item: *Gnome eats pizza at three o'clock.*

Meadow *et al.* (2008)).

In the pictures, temporal information is depicted by a clock showing a time; see the example in figure 6.1. Participants thus only had to communicate about events taking place at a certain *time*. This is the kind of temporal information that turned out to be the easiest to convey in the pilot study.

6.3.2 Expressing temporal information in Dutch

Participants were first asked to describe each picture using a Dutch sentence, and subsequently, they were asked to gesture. Because the structure of Dutch sentences with temporal information is not exactly parallel to English, let me describe the orders that are acceptable for normal spoken Dutch, when, e.g. figure 6.1 is described. In English, the two possible orders are 'A gnome eats pizza at three o'clock' and 'At three o'clock, a gnome eats pizza'. In Dutch, there are three possibilities (the order of the sentences is indicated below each example):

(3) Om drie uur eet de kabouter pizza.
At three o'clock eats the gnome pizza.
(TVSO)

(4) De kabouter eet om drie uur pizza.
The gnome eats at three o'clock pizza.
(SVTO)

(5) De kabouter eet pizza om drie uur.
The gnome eats pizza at three o'clock.
(SVOT)

Thus, the temporal information (T) can be put at the beginning (3), the end (5), or in the middle, between the verb and the direct object (4). Note that when the adverb is fronted (like in (3)), the word order of the rest of the sentence changes. This occurs because Dutch is a V2 language: the verb is always the second constituent of a declarative sentence. Dutch behaves like an SVO language for simple expressions, such as the following example:

(6) Een kabouter eet pizza.
A gnome eats pizza.
(SVO)

In this example, the verb is the second constituent of the sentence. But when a temporal adverb is the first constituent of the sentence, such as in example (3), the word order of the rest of the sentence inverts in order to keep the verb in second position.

To sum up, in spoken Dutch three orders are possible to express the information in the picture, and each order has the temporal adverb in a different place: start, middle or end. Moreover, the word order of subject and verb is influenced by the placement of the temporal adverb: if the adverb is fronted, the order of the rest of the sentence is Verb-Subject-Object (VSO).

6.3.3 Aims and predictions

Word order in speech and gesture

In chapter 5 we have seen that when people communicate about simple events in an improvised manner, they by-pass the grammatical rules of their native language. If this behaviour extends to more complex expressions, we expect to see no occurrences of the TVSO structure that is particular for Dutch grammar. Alternatively, if more complex utterances would trigger participants to start using grammar rules, such as was suggested by Langus and Nespor (2010),[1] we would indeed observe TVSO gesturing sequences.[2]

[1]See the previous chapter for more details about the studies carried out in Langus and Nespor (2010).

[2]In this chapter, I will use 'S,' 'V,' and 'O' instead of 'Ar,' 'P,' and 'A' to refer to gesturing sequences. I am aware that the latter are more suitable to describe sequences in a system that is governed by semantic principles and not by syntactic rules, but I will use S, V and O because they increase readability.

First of all, we compared the participants' placement of the temporal adverb in speech and gesture. In spoken Dutch, one is free to use any of the three word orders mentioned above. In gesture, if participants indeed apply the temporal displacement strategy, we expect participants to be more likely to put the temporal adverb in a frontal position than in speech.

Secondly, we tested whether the TVSO word order, that is specific for Dutch grammar, is found in the gesturing orders of the participants. As we have seen above, when the temporal adverb is fronted, Dutch grammar prescribes a change in the order of the rest of the sentence: this should then become VSO order. If participants are indeed bypassing the rules of their native language, we expect the TVSO order to occur less often in the gesturing strings than in the speech strings.

Influence of order of presentation

To test whether the order in which the information on the picture is presented has an influence on gesturing order, we presented the stimuli in two ways. On each picture there is temporal information (the clock time in the upper right corner) and propositional information (the event displayed in the center of the picture). When showing a picture to a participant, we cannot be sure which information is looked at first. Therefore, we presented the information in stages (see figure 6.2): for some pictures, the temporal information was shown first, and the event information was added later. In other pictures, the event information was shown first, and the temporal information was added later. If the order of presentation is relevant for the participants, it will show in the results.

Intensional and extensional events

Finally, we tested the interaction between the gesturing of temporal information versus propositional information (or event information). The events used in the experiment were either intensional (e.g., 'pirate thinks of shopping cart' in figure 6.2) or extensional events (e.g., 'gnome eats pizza' in figure 6.1). In the previous chapter, it was shown that when participants communicate about these two kinds of events in an improvised manner, they prefer to use SVO order for intensional events, and SOV order for extensional events. If the improvised communication task indeed makes participants bypass the grammar rules of their native language, then we expect this preference to remain the same when temporal information is added to the events.

6.3.4 Method

Participants

Sixteen participants (9 male, 7 female; age range 17–26) were recruited from Utrecht University. All participants were native speakers of Dutch; none of them

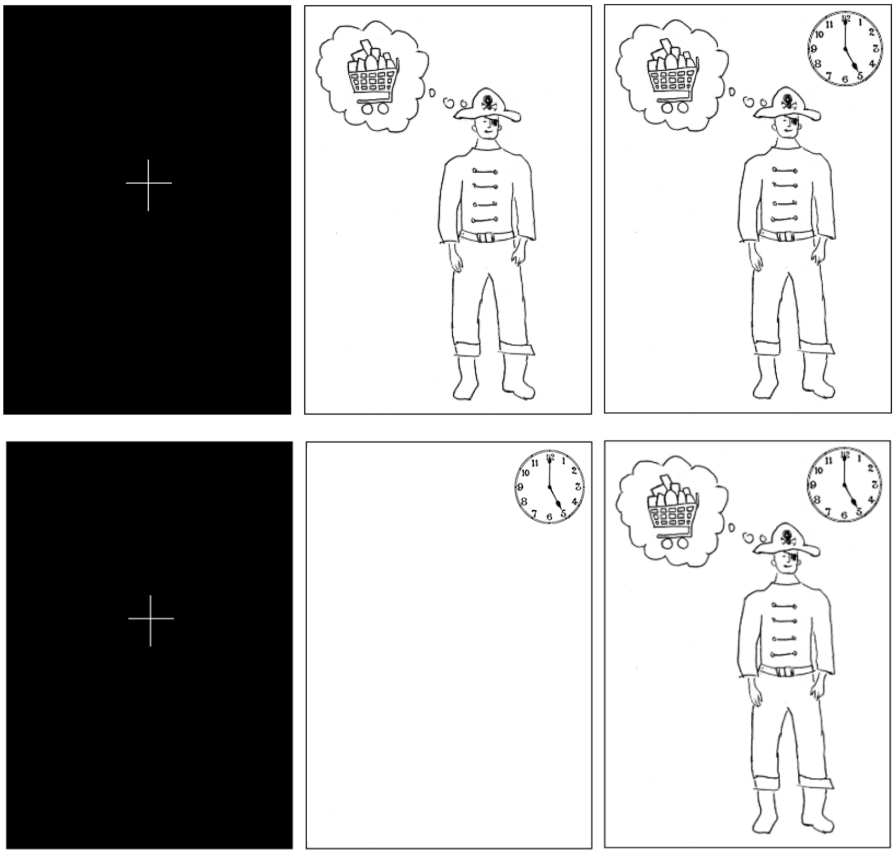


Figure 6.2: These examples show the two orders in which the information was presented in the experiment. At the top, (1) a black screen with a fixation cross, (2) the event information, (3) the event information plus the temporal information. At the bottom, (1) a black screen with a fixation cross, (2) temporal information, (3) the event information plus the temporal information.

were familiar with any conventional sign language. The participants received a monetary compensation for participating.

Stimuli

16 pictures of simple events with a picture of a clock indicating a time between 1 and 10 o'clock in the upper right corner were used as stimuli. Of the events in the pictures, 8 were extensional events (motion events) and 8 were intensional events (see chapter 5). Pictures were presented in random order. There were two practice items (one extensional and one intensional).

Procedure

The experiment consisted of two parts. In part one, participants were shown the items on paper and were asked, for each item, to describe the information presented on the picture in Dutch. Spoken sentences were recorded on the laptop using Audacity.

In the second part, participants were told that they would get to see the same pictures again, but now they should convey the information on the pictures using only gesture and no speech. The pictures were shown on a screen and the information on each picture was presented in stages. In eight items, the event information was shown first, and the temporal information was added afterwards (as in figure 6.2). In the eight remaining items, the temporal information was shown first, and the event information was added afterwards. Participants could use the mouse button to click through the two stages, in order to see the complete scene. Between the two stages, a delay of at least 2 seconds was built in, in order to make sure that the participants would really see the information in stages.

Two practice items were used for both the speech part and the gesturing part of the experiment. During the gesturing part, the experimenter gave feedback about whether all the information on the picture was conveyed. During the actual experiment, the experimenter was sitting behind the screen and did not give any spoken feedback. The participants were filmed using the built-in webcam of the laptop.

6.3.5 Data analysis

The speech strings from the first part of the experiment were transcribed and coded for order. The videos were analyzed by two independent coders. The items for which the participant had started to gesture before all the information (the temporal information as well as the event information) was visible, were excluded (5 of 256 items). The remaining 251 items were coded for gesturing order by two independent coders, with 94.8% agreement. Items for which there was no agreement were excluded (13 of 251 items).

order	**# items**	**%**
SVOT	128	50.0
TVSO	39	15.2
SOVT	37	14.5
SVTO	21	8.2
STOV	13	5.1
other	18	7.0
total	256	100

Table 6.2: Speech orders used by the participants and their frequency.

order	**# items**	**%**
TSVO	104	43.7
SVOT	47	19.7
TSOV	36	15.1
SOVT	33	13.9
TVSO	2	1.3
other	16	6.3
total	238	100

Table 6.3: Gesture orders used by the participants and their frequency.

Let me give a brief overview of the orders that were found in the speech part and the gesturing part of the experiment. This is just to provide a very general overview; the results of the statistical tests will be discussed in the next section.

When speaking, participants used more different orderings than expected on the basis of the three example sentences above (sentences (3), (4) and (5)). This is because some participants formulated sentences like the following.

(7) Een kabouter die pizza eet om drie uur.
A gnome that eats pizza at three o'clock.
(SOVT)

(8) Een kabouter die om drie uur pizza eet.
A gnome that eats pizza at three o'clock.
(STOV)

Thus, for the spoken descriptions of the items, 5 different orders were found; see table 6.2 for an overview of these orders and their frequency. The category 'other' consists of those sentences where people repeated information, or omitted information in their description. When gesturing, on the other hand, participants were completely free to choose any order they wanted, but still, there were only four dominant orders; see table 6.3 for these orders and their frequency. The category 'other' in that table represents those orders where participants repeated gestures or otherwise used an order different from the four dominant orders.

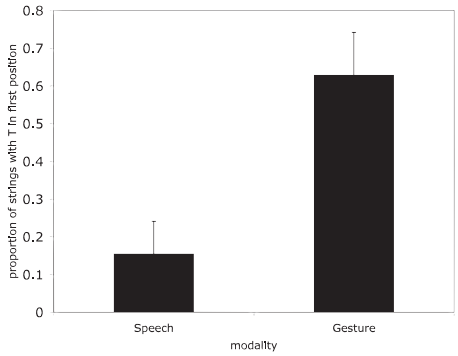


Figure 6.3: Mean proportions of strings with the temporal adverb in initial position in speech and gesture. Error bars indicate standard error of the mean.

	T first	TVSO	% TVSO
speech	39	39	100%
gesture	158	2	1.3%

Table 6.4: Occurrence of VSO order in strings where T is first, in speech and gesture.

6.3.6 Results

Placement of T in speech and gesture

The placement of the temporal information in the speech strings was compared to that in gesture strings. A repeated measures ANOVA shows significant main effect for modality on the placement of the temporal adverb $F(1, 15) = 15.940, p = .001$. Pairwise Bonferroni corrected comparisons show that the temporal adverb was placed in the initial position more often when participants were gesturing ($M = .629, SE = .113$) than when they were speaking ($M = .155, SE = .086$). See figure 6.3.

Occurrence of VSO order

The inversion of word order after fronting of the temporal adverb in speech was compared to that in gesture, see table 6.4. In 39 (of 256) *speech* strings, the temporal adverb was placed in the initial position, and in all of these, the word order was TVSO. Of the *gesture* strings, 158 (of 238) had the temporal adverb in initial position, and in only 2 occurrences was the specific order TVSO (the dominant orders were TSOV and TSVO).

There is a significant difference between the two modalities with respect to the occurrence of TVSO order in those strings where T is put in initial position $\chi^2(1)$=179.076, $p < .0001$.

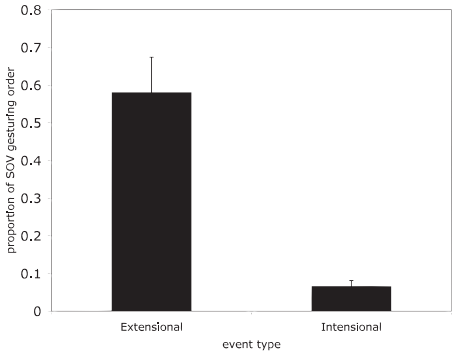


Figure 6.4: Mean proportions of strings with SOV order for extensional and intensional events. Error bars indicate standard error of the mean.

Influence of order of presentation

It was verified whether the order of presentation of the information (temporal information and event information) in the items had an influence on the placement of the temporal adverb: did participants change the placement of the temporal adverb in their gesturing if the information was presented in a different order? Using a repeated measures ANOVA, we found no significant main effect of order of presentation of the information (temporal information first vs. event information first) in the pictures on the placement of the temporal adverb $F(1, 15)$=1.931, p=.185.

Intensional and extensional events

It was investigated whether the difference between intensional and extensional events would lead to different orderings. A repeated measures ANOVA shows a significant interaction between the gesturing order of the event information and the kind of event (intensional or extensional). $F(1, 14)$=37.062, p=.000. Pairwise Bonferroni corrected comparisons show (see figure 6.4) that participants use SOV order more often for extensional events (M=.581, SE=.093) than for intensional events (M=.066, SE=.015).

6.4 Discussion

6.4.1 The temporal displacement strategy in the lab

The study aimed to investigate what happens when participants in an improvised communication task are asked to convey complex propositions. It was investigated, first of all, whether the temporal displacement strategy observed in adult homesign and the Basic Variety would be replicated in the lab and,

secondly, how the temporal and the propositional information in the gesturing sequences of the participants interacted.

The placement of the temporal information in the gesture strings was compared to two things. First of all, to the placement of the temporal adverb in the spoken sentence. Secondly, to the order of presentation of the information in the items (in each item the information was presented in stages; see the description in section 6.3.3).

Participants were expected to place the temporal information at the beginning of the gesturing string more often than in the speech string, because for improvised communication, the only known way to indicate that a certain event takes place at a certain time is by applying the temporal displacement strategy. This hypothesis was confirmed.

In the gesturing part of the experiment, each item was presented in stages. In eight of the 16 items, the temporal information was presented first, and in the remaining eight items, the event information was presented first. One could expect that this order matters to participants, because in the first case, the information is presented like this (I am using the example that was used in figure 6.2):

(9) There is a time (five o'clock). At this time a pirate thinks of a shopping cart.

In the second case, the information is presented like this:

(10) There is a pirate who thinks of a shopping cart. This takes place at five o'clock.

Put differently, only in the case of example (9) does the temporal information introduce a temporal frame; in the other example, the temporal information is focus information. Despite the fact that the items in the experiment were presented in these two ways, this had no effect on the order of the gesture strings. Thus, the participants were not influenced by the order of presentation. This suggests that in improvised communication, the temporal displacement strategy was not influenced by the informational status that was given to the elements in the presentation of the items. In other words, even when temporal information was suggested to be focus information in the presentation of an item, participants used it to define a temporal frame for the remaining information (see chapter 4).

We can thus conclude that the temporal displacement strategy that was found in existing restricted linguistic systems (as described in chapter 4) has been reproduced in a lab situation. This strengthens the claim put forward in chapter 4 that we can use the improvised communication experiment to collect restricted linguistic data. Because a lab experiment gives us more control than natural situations, the improvised communication task allows us to look in

more detail at the principles that play a role when people construct utterances in an improvised manner. In the experiment we looked in more detail at the structure of the gesturing strings that were produced to investigate how the temporal information and the propositional information influenced each other.

6.4.2 The structure of complex improvised utterances

To investigate how the temporal and the propositional information interact in improvised communication, we studied two issues. First of all, we compared the occurrence of an order that is specific for Dutch grammar in the spoken sentences to that in the gesturing strings. Because Dutch is a V2 language, whenever a temporal adverb is put in first position, the remainder of the sentence is ordered differently (namely, as VSO). It was found that in improvised communication, this order is hardly ever used (it was used in only 2 cases).

Secondly, we investigated whether people would apply the same orderings that were found in the previous chapter, when communicating about extensional and intensional events. We found that, despite the addition of temporal information, participants still typically gestured extensional events in ArPA order and intensional events in ArAP order. In other words, the strategy that was observed in chapter 5 to use different orders to convey information about simple events, depending on whether these events are extensional or intensional in nature, remains visible even when information is added and the message becomes more complex.

We can conclude from this that people use different strategies when they are engaged in improvised communication than when they speak their native languages, and that they bypass the rules of their native languages. This was already shown for simple (intensional and extensional) events in the previous chapter and in previous publications by Goldin-Meadow *et al.* (2008) and Langus and Nespor (2010), but it is thus confirmed for more complex events in this chapter (contra Langus and Nespor (2010)).

It was found in the study presented in this chapter that when a simple proposition is combined with information about its temporal location, these two kinds of information are combined in one utterance, but they seem to remain separate to a certain extent:

- The temporal information (T) is expressed before the propositional information (TSVO and TSOV gesturing orders), or sometimes after it (SVOT and SOVT gesturing orders), but it never breaks it up (see table 6.3).

- The fact that temporal information is expressed does not interfere with the ordering patterns found previously: SOV order is still preferred for extensional events and SVO order is preferred for intensional events.

It looks like the information in these complex utterances consists of two separate parts: a core part, in which information about the simple proposition

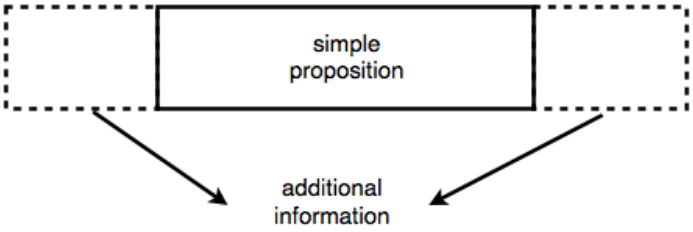


Figure 6.5: Partitioned organisation of a complex expression: simple propositional information embedded in additional information.

is expressed, and the periphery (the two outer edges), in which information about its temporal location is expressed, as depicted in figure 6.5.

In figure 6.5, it is the simple proposition that forms the core of a complex expression. In this simple proposition, the organisational principles are of a semantic nature: they tell us how the verb and its arguments should be organised.[3] When temporal information is added to a simple proposition, this is added to one of the edges of an utterance: most often the beginning and alternatively the end of the utterance. In a way, it seems logical that the temporal adverb is not mixed with the information about the event: the temporal adverb determines the temporal location *of the whole event*.

By contrast, looking at full-fledged languages, we see that temporal information is given mainly through inflection of the verb.[4] The temporal information has moved into the 'core' part of an utterance and has become part of it. The comparison of improvised communication with full language reveals that temporal information has moved from the edges of an utterance *into* the utterance.

This shift from the edges to the inside of an utterance can be observed in other domains, for example that of negation. In de Swart (2009), the expression of negation in varieties of language learners is described. The paper compares the Prebasic Variety to the Basic Variety. The Prebasic Variety is a stage in the learning process of second language learners who learn a language outside the classroom, and it precedes the Basic Variety stage. When negative information is expressed in this stage, the negator appears at one of the edges of an utterance. See these examples from (de Swart, 2009, p. 66):

(11) veel eten nee
much eat no

[3]AgentFirst is one example of such a principle; it tells the speaker or hearer that the argument with greatest control comes first. The semantic properties of the event play a role at this level too: extensional events first need all arguments, before the action itself is described (Actor-Patient-Act), and intensional events are organised according to the principle Actor-Act-Patient.

[4]Of course, temporal adverbs are also used in full languages, but they are optional, whereas verb inflection is obligatory for the majority of languages.

(12) nein tasche eh links
no bag eh to the left

For negative information, like for the temporal information, it is in some sense logical to put the negator at one of the edges of an utterance: it is the proposition as a whole that is negated. Still, when learners reach the Basic Variety stage, they start to embed the negator in the sentence, such as in this example from (de Swart, 2009, p. 69):

(13) ik niet ∗hapis∗ gaan
I not prison go

Thus, as the language becomes more sophisticated, negation moves from the edges of an utterance into the core of the utterance. To draw a general conclusion from these observations, it appears that the complexification of language followed a route in which information was first added to the outer edges of simple utterances, and later absorbed into the core of an utterance with the further development of language.

I speculate that the process we have observed in this section, of information moving from the edges to the core of an utterance, co-occurs with the growing influence of newly emerging conventions. This can be purely an effect of repeated interaction among language users, or it can be driven by the need to formulate more precise messages. This is only a speculation, and the experiment in this chapter does not offer us any evidence on how and why this pattern would occur, but it can certainly be tested empirically. In the next chapter, I will briefly discuss ways to investigate this possibility.

6.5 Conclusion

In this chapter I investigated what happens when we ask people to convey complex information, that is, propositional information plus information about the time at which the proposition is true. It turned out that when people are asked to communicate in an improvised manner, there is a general preference to convey information about the time of the event at the beginning of an utterance. They use the beginning of the utterance to sketch the circumstances under which the event described in the remainder of the utterance takes place. This finding is in accordance with the displacement strategy observed in existing restricted linguistic systems like adult homesign and Basic Variety. The fact that we have replicated the displacement strategy in the lab supports the claim that the improvised communication task can be used in addition to existing restricted linguistic systems, to gain indirect evidence about the emergence of syntactic structure.

When looking in detail at the strings that were produced in the study presented in this chapter, we have seen that the expression of information about the time of an event does not influence the way in which the rest of the

utterance is expressed. Temporal information is expressed at one of the edges of an utterance, and the internal structure of the part that expresses a simple proposition does not change under the influence of the presence of temporal information. By contrast, in full-fledged languages, temporal information is incorporated in the verb, as it is typically expressed through verbal inflection. I have thus specified a possible pattern according to which language could have become more complex: information added to a simple proposition is introduced at one of the edges of an utterance in simple language systems, and becomes part of the core of an utterance in more complex language systems, including full language.

CHAPTER 7

Conclusion

In this chapter I review the conclusions and results from the chapters of my dissertation. I will start out by providing an overview of all the chapters in this dissertation, and subsequently I will reflect on these results. Lastly, I will sketch directions for future research.

7.1 Overview of the chapters

7.1.1 Semantic protolanguage and restricted linguistic systems

Meaning deserves more attention than it gets in the language evolution debate. Instead of focusing solely on possible steps in evolutionary history that have led to structural complexity of language, we should take meaning into account as well. Not only meanings of individual words, but meaning as a compositional phenomenon. Human language is compositional, and this means that complex meaningful structures can be analysed in terms of the parts they consist of, and the way in which these parts were put together. In this way, humans can build complex meanings from simpler ones. Compositionality could have started very simple, in a stage where only few meaningful items were put together. This intermediate step would then bridge the gap between simple communication and full syntactic language. In other words, an account of the emergence of compositionality might give us the emergence of syntactic complexity 'for free'.

Currently, many scholars agree that the step from non-language to full language was not one giant leap, but there have been one or more intermediate

steps. One way to talk about these intermediate stages in the evolutionary history of language is to use the notion of protolanguage. Many different accounts of protolanguage have been provided in the literature, and I review some of them in section 1.4. I conclude that different accounts of protolanguage stress different mechanisms that could have played a role in the emergence of language. These different accounts are not necessarily mutually exclusive: many processes might have played a role at the same time.

A place where meaning and protolanguage come together is in the **semantic protolanguage** account. It is an account of protolanguage based on the idea put forward by Jackendoff (2002) that there has been a stage in which language was governed by semantic (or pragmatic) principles like AgentFirst and FocusLast. If this stage has indeed existed, its organisational principles can be seen as a predecessor of fully syntactical language rules. In chapter 1, I formulate the main questions for my dissertation:

- Is there further evidence for a semantic protolanguage?
- What are the mechanisms in semantic protolanguage?

In chapter 2, I give an overview of ways to collect empirical evidence about the emergence of language. Because we do not have any direct evidence about the beginnings of language, the evidence that we resort to is *indirect* evidence. One of the ways to collect indirect evidence is the study of so-called *restricted linguistic systems*. From a discussion of these systems, the general picture appears that when people cannot use or learn language in a normal way, they still organise their utterances in a semi-systematic way, using principles that do indeed seem semantic or pragmatic in nature.

Some problems with this approach remain, however. One of these problems is that in order to draw more well founded evolutionary conclusions, we need to specify the principles on the basis of which we can draw these evolutionary conclusions (a *bridge theory*). Secondly, it would be good to be able to collect more data, in a more controlled manner, i.e., in a lab situation instead of in natural settings.

7.1.2 Meaning and evolution

Before the issues brought forward in chapter 2 are tackled, I focus in more detail on meaning in chapter 3. The answer to a question like 'what is meaning?' depends on the setting in which the question is asked. In philosophy, many scholars have thought about the nature of meaning without taking its evolution into account. The different characterisations these scholars have come up with are useful, because they show what is important about meaning in full language. From an overview of accounts of meaning, I distill three intuitions:

reference The things words and sentences refer to are important when we want to know their meaning.

intention and belief When a speaker makes an utterance, he typically intends his audience to come to believe something.

convention Our utterances have the meanings they have because these utterances have been used by other speakers in similar situations.

These intuitions can be seen as competing definitions of meaning, but they can also be seen as different ways to shed light on a multi-faceted phenomenon.

When meaning is studied from an evolutionary perspective, different demands are placed on the definition of meaning. Once we look at the emergence of meaning, we cannot simply assume that there are abstract objects called meanings (such as is done in propositional accounts of meaning). I focus on two possible trajectories from animal behaviour to full linguistic behaviour in humans: one sketched by Peter Gärdenfors, the other by Ruth Millikan.

Peter Gärdenfors postulates that cognition develops before meaning. In his view, private mental representations constitute the basis for sophisticated communication. This view implies that apes, our closest cousins, may have well developed cognitive skills, but simply do not communicate in a similarly sophisticated fashion. Other authors who wrote about the emergence of meaning in human language take this position as a starting point (Fitch, 2010; Hurford, 2007). Ruth Millikan focuses less on cognition. She emphasises the conventional properties of language, and compares language to other cultural phenomena: behavioural patterns which are copied by individuals, like the habit of shaking hands when meeting in many western societies.

There is much to say for a Gärdenfors-like approach to the emergence of meaning, which emphasises the relation between cognitive structures in language users, and their utterances, because it points our attention to cognitive properties of our ancestors (animals). Their cognitive capacities are ancestors of features of our linguistic capacities. Much interesting research has been carried out already to describe these links (Fitch, 2010).

The line of thinking introduced by Ruth Millikan points our attention away from cognitive structures to conventions, to language as a public phenomenon. Millikan denies that there is a one to one relationship between the meaning of an utterance and what is mentally represented by the speaker or the hearer. Emphasising the conventional nature of linguistic meaning and grounding conventional behaviour biologically, she takes the communicative nature of language into account.

Gärdenfors's and Millikan's views represent two ways of thinking about the evolutionary trajectory that language took: a cognitive and a communicative

road to meaning in full language.

A cognitive road to meaning: an evolutionary scenario in which cognitive capacities are central, and are held responsible for the structure of the emerging language.

A communicative road to meaning: an evolutionary scenario in which the dynamics of communication and the formation of conventions are central.

These two trajectories represent different emphases in thinking about the emergence of meaning: one that stresses private, cognitive processes in individuals, and one that stresses public, communicative processes. I have claimed that in order to arrive at a balanced picture of the emergence of language, both sides need to be taken into account.

7.1.3 Cognition and communication in the wild and in the lab

Armed with these principles and insights about meaning and its emergence I focus on semantic protolanguage again in chapter 4. Data from restricted linguistic systems can be used as evidence for the existence of a semantic protolanguage stage, but some problems must be addressed: more data is necessary in order to make a semantic protolanguage account well founded and detailed enough to formulate precise hypotheses about the actual principles that have played a role in semantic protolanguage.

To obtain more empirical data, I propose to conduct lab experiments in which participants are forced to communicate in an improvised manner by using only gesture and no speech.

In order to formulate well founded evolutionary claims on the basis of restricted linguistic systems, we need a proper *bridge theory*. A bridging theory takes us from empirical data to conclusions about protolanguage, and clarifies how the analogy between restricted linguistic systems and protolanguage works.

In recent literature, two such accounts have been provided. There is a bridge theory in which the analogy between restricted systems and protolanguage is claimed to be *cognitive*, and another one in which the analogy rests on *communicative pressures*. According to the first, there is an analogy between the cognitive structures in speakers of restricted language and speakers of protolanguage. The assumption behind this is that cognitive structures are directly responsible for the structure of utterances. According to the second approach, there is an analogy between the situation of protolanguage speakers and speakers of restricted language. The utterance structures that show up in restricted linguistic

systems are similar to protolanguage, it is claimed, because the communicative pressures are similar. Thus, this bridge theory makes use of the assumption that communicative pressures are responsible for the structure of utterances.

I provide an analysis in which I first show that different restricted systems emerge in slightly different circumstances, and these circumstances might influence their structure. For that reason, I show, a full-fledged bridge theory has to take into account both cognitive and communicative aspects. Restricted linguistic systems emerge in situations where there are newly emerging linguistic conventions. Because these conventions are not fully in place, the way they are in normal language, speakers and hearers have to improvise when constructing and interpreting utterances. When speakers in such situations put an utterance together without having a full system of conventions to fall back on, they may have cognitive biases that influence the utterance structures they choose. But it is not necessarily the case that the utterance structures that are found in restricted linguistic systems are direct mirror images of the mental representations of the speakers. I imagine that the formation of utterances in restricted linguistic systems is an active process that goes from multidimensional mental representations to linear structures.

This process of linearisation of information is (1) an active process, and (2) something inherently communicative. It is active, because every time an individual wants to communicate propositional information, it has to be translated into a string. It is essentially communicative, because the reason that a mental representation is translated into a string is that it is communicated. My prediction is that in the absence of stable linguistic conventions, communicating individuals will linearise information differently when this serves a communicative purpose.

Thus, the key to my bridge theory is the similarity of the situation individuals (in the improvised communication setting and in prehistoric times) find themselves in: they have to linearise non-linear mental representations into a string, in the absence of a stable system of linguistic conventions. The absence of linguistic conventions forces people to bypass the way in which they would normally structure information and build sentences (according to the rules of their own language), and I speculate that it triggers an evolutionarily old strategy: to structure information according to its semantic properties. This speculation would imply that the old strategy has been 'overruled' by newer strategies (syntactic structure) in the course of evolutionary history, and this connects to ideas expressed elsewhere in the language evolution debate, such as the ones expressed in Hurford (2011); Tomasello (2001).

7.1.4 Improvised communication

In the final chapters I present improvised communication studies, looking at the expression of simple propositions, and more complex propositions. In chapter 5, I start from the observation made in Goldin-Meadow *et al.* (2008) that Actor-Patient-Act (or SOV) is the order in which simple events are represented

mentally. I show that besides the extensional (motion) events used in Goldin-Meadow *et al.* (2008), a second kind of event must be distinguished, namely intensional events. These events have different semantic properties. Whereas for extensional events, there is simply an Actor, a Patient and an interaction between the two in an Act (for motion events, the Act involves motion), for intensional events, the Patient is more abstract, because it either does not exist, is not specific, or otherwise is dependent on the Act.

In an improvised communication study, I show that the different semantic properties of these two kinds of events lead to two different gesturing orders: Actor-Patient-Act (SOV) for extensional events and Actor-Act-Patient (SVO) for intensional events. This confirms my prediction that different semantic properties of events will make speakers change the way they linearise information, when they communicate in the absence of an existing system of conventions.

In order to show that this linearising behaviour is there for a communicative purpose, I describe a second study, in which participants are asked to interpret improvised behaviour. Using videos with ambiguous actions, I show that depending on the order in which the elements in a string are presented (SVO or SOV), people choose different interpretations (intensional or extensional ones). The fact that different orderings lead to different interpretations confirms that the different orderings have a communicative purpose.

Chapter 6 investigates how more complex information is expressed in improvised communication. In other words, what happens when information is added to a simple proposition? How is this communicated in a simple language system?

The chapter describes an improvised communication experiment in which participants were asked to convey, not only information about simple events, but also about the time at which the event took place. In existing restricted linguistic systems such as adult homesign and the Basic Variety, it has been observed that people can use relatively simple means to convey that a certain event does not take place *now*, but rather in the past or the future. They do this by applying the temporal displacement strategy: by placing a temporal adverb at the beginning of an utterance, the event described can be located on the time axis at some other moment than *now*. Participants in the improvised communication study were found to replicate this strategy. Moreover, it was found that in the utterances observed in the experiment, simple propositional information is never interrupted by temporal information, because temporal information is always expressed at one of the edges. Finally, the principles that govern the transmission of simple propositional information (such as the distinction between extensional and intensional events observed in chapter 5) were not influenced by the presence of temporal information. This suggests that simple propositional information and temporal information make up two separate partitions. The propositional information forms the core of an utterance and temporal information is added at the outer edges.

7.2 Answers to the central questions

The goal of this dissertation was to explore the semantic account of protolanguage by answering the two questions formulated above: Is there further evidence for a semantic protolanguage? What are the mechanisms in semantic protolanguage? I will formulate these answers here.

7.2.1 Further evidence for semantic protolanguage

Jackendoff (2002) hypothesised that there has been a stage in the emergence of language in which multiple words were concatenated according to semantic and pragmatic principles rather than syntactic rules. These semantic principles are less strict than the syntactic rules we have in full language, and they derive directly from the semantic properties of the words they apply to. Examples of such principles are AgentFirst (when describing a situation, mention the individual with the highest control first) and FocusLast (information that is in focus should be at the end of an utterance). Evidence for the existence of such a protolanguage stage has been derived from language systems that emerge in situations where people cannot use or learn an existing language. These systems are called restricted linguistic systems, and examples are pidgin, Basic Variety and newly emerging sign languages.

In this dissertation, I have added a source of evidence to this list: the improvised communication experiment. When participants in such an experiment are asked to communicate using gesture and no speech, they have to improvise, using an inventory of highly iconic gestures. Previous experiments had already shown that people in such an experiment bypass the grammatical rules of their native language. The experiment can thus be seen as an environment in which the beginnings of a newly emerging language system are made.

I have argued that gesture strings observed in the improvised communication task can tell us something about protolanguage because they put people in a situation that is similar to that of our protolanguage speaking ancestors. There is no full system of linguistic conventions, and when speakers make utterances, the organisation of these utterances will depend on how a speaker goes from a non-linear mental representation to a string (of sounds or gestures). The way in which information is linearised into strings (utterances) is thus dependent on the cognitive biases of the speaker, but it is at the same time a communicative phenomenon, because the linearisation of information is something that is done *for communication*.

7.2.2 The mechanisms in semantic protolanguage

I have shown that when people are engaged in improvised communication, the organisation of their utterances will depend on the kind of event they describe. Besides the class of extensional events, events in which *someone (Actor) does something (Act) to someone or something (Patient)*, I have defined the class of

intensional events. In intensional events, the Patient is more abstract than in motion events, because it either does not exist, is not specific, or is dependent on the Act. When communicating about intensional events, participants in the experiment deviate from the Actor-Patient-Act order that was previously found for extensional events (Goldin-Meadow *et al.*, 2008), and use Actor-Act-Patient. This behaviour is independent from their native language. An interpretation study reveals that the two different orders lead to different interpretations, and thus, the claim that variation in word order is applied for communicative reasons is confirmed.

When people engaged in improvised communication go beyond simple propositions and add information about the time of an event, they prefer to put this at the beginning of an utterance. This was first observed in existing restricted systems (Benazzo, 2009), but I have replicated this pattern in an improvised communication study. This study further reveals that the presence of temporal information does not influence the way in which the propositional information is expressed. The two kinds of information seem to operate on different levels.

7.3 Cognition and convention in newly emerging languages: directions for future research

In this dissertation, situations in which new language systems develop are depicted as follows. In situations without a full blown language, there is no system of grammar rules that tells a speaker how to organise his utterances. When a speaker wants to convey a thought, he is forced to linearise this thought, because language (spoken as well as signed) forces us to use string-like utterances. The way in which a communicating individual linearises information is influenced by cognitive biases. Once an utterance has been produced, it has made its way out into the world: a hearer interprets the utterance, and it is possibly repeated, depending on its communicative success. As soon as a certain structure is repeated multiple times, conventions might be formed about the organisation of utterances, and these newly formed conventions might become the grammar rules of the newly formed language.

Thus, in newly emerging languages, *cognition* initially plays a role in the communicating individual, and determines how information is linearised into a string. *Communicative pressures* start to play a role after utterances are used among individuals, and constructions are repeated or discarded, according to their communicative success (see figure 7.1).

The improvised communication experiments I conducted focus on gesture production and interpretation separately, using only one speaker or one interpreter. In such a setting, newly emerging conventions play a marginal role, because there is no chance for communicative pressures to have an effect.

It would be an interesting direction of research to investigate what happens once we do allow communicative pressures among communicating individuals

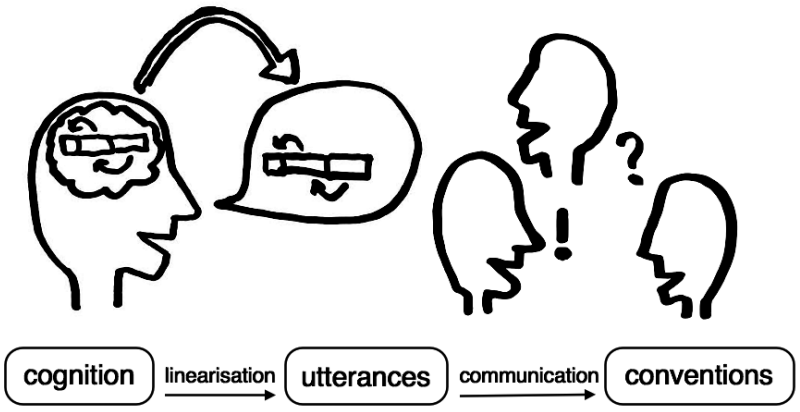


Figure 7.1: From cognition, via linearisation to utterances, and from single utterances, via communicative pressures to linguistic conventions.

to play a role. This would be possible if we would conduct the improvised communication experiment in a **repeated interaction** setting: participants are asked to produce and interpret gesturing sequences alternately, in several consecutive rounds. Previous research, with graphical communication tasks, has shown that in an interactive setting, less detailed (more symbolic and less iconic) and more effective graphical signs are produced after a series of interactions between participants (Garrod *et al.*, 2007). This can be seen as the emergence of a shared set of conventions.

In repeated interaction experiments with improvised gestured communication, I would like to investigate the influence of the emergence of conventions on gesturing patterns. The experiment will focus on the distinction between intensional and extensional events. If these two kinds of events are used in an interactive improvised communication experiment, will this result in two different orderings (similar to what was shown in chapter 5), or will one of the two orderings disappear after several rounds of interaction? Together, improvised communication and social transmission can provide a clearer picture of how cognitive and communicative biases come together to give us linguistic structure.

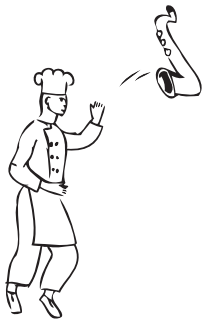
This is only the beginning: we are still far from understanding where human language came from, and this dissertation has only laid the bare basis of an approach to investigate the emergence of language which takes meaning seriously. But it does show that doing experiments to study the origins of language is indeed fruitful and offers many possibilities for future research.

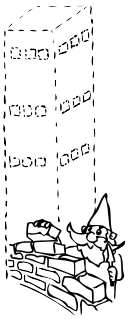
Appendix A: Stimuli Chapter 5

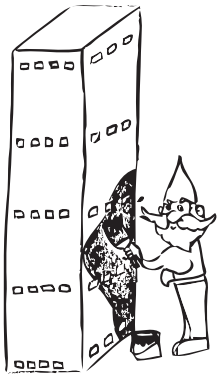
Production experiment

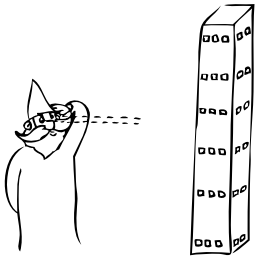
The following is a list of descriptions of the items used in the production experiment. Test items were 'gnome pushes cart,' 'princess dreams of car,' 'princess throws sax,' 'witch pulls car,' and 'pirate thinks of cart.' All images are provided below.

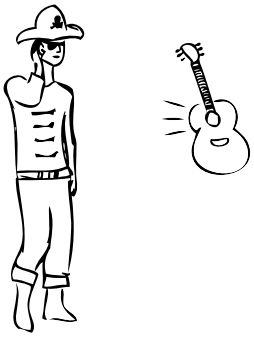
version A	**version B**
cook cuts sock	princess cuts scarf
cook knits sock	princess knits scarf
cook swings saxophone	pirate swings guitar
cook throws saxophone	pirate throws guitar
gnome builds tower	witch builds house
gnome climbs tower	witch climbs house
gnome eats pizza	witch eats banana
gnome wants pizza	witch wants banana
pirate carries ball	princess carries vase
pirate looks for ball	princess looks for vase
pirate dreams of guitar	cook dreams of saxophone
pirate hears guitar	cook hears saxophone
princess drops vase	pirate drops ball
princess sculpts vase	pirate sculpts ball
princess hangs scarf	cook hangs sock
princess thinks of scarf	cook thinks of sock
witch paints house	gnome paints tower
witch sees house	gnome sees tower
witch cuts banana	gnome cuts pizza
witch draws banana	gnome draws pizza

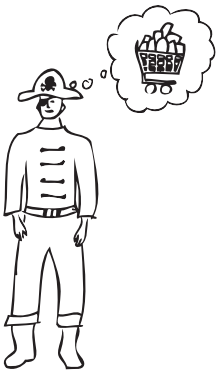


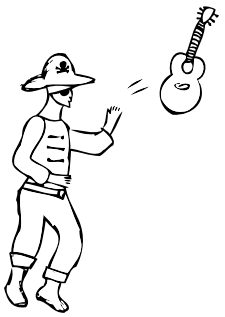


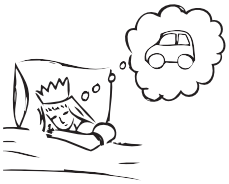


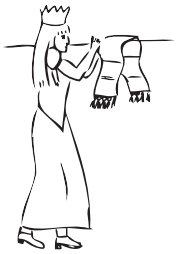


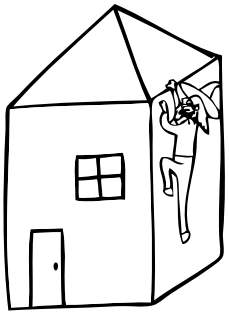












Interpretation experiment

The three images below represent the three stages for each item of the interpretation experiment: first, the two answer possibilities were shown, subsequently, the video was shown, and finally, the two answer possibilities were shown again, and the participant was asked to make a choice.

The following is a list of descriptions of the videos used in the experiment. Each video was played either in ArAP or ArPA order (depending on the version; each version had 6 videos in ArAP order and 6 in ArPA order), and the answer possibilities were always the two possible interpretations of the action, presented as images in random order.

- Pirate drops$_m$/searches$_i$ ball.
- Princess breaks$_m$/sculpts$_i$ vase.
- Leprechaun cuts$_m$/draws$_i$ pizza.
- Witch eats$_m$/wants$_i$ banana.
- Witch paints$_m$/paints$_i$ table.[1]
- Girl sleeps on$_m$/dreams of$_i$ book.
- Girl kisses$_m$/thinks of$_i$ doll.
- Princess talks to$_m$/talks about$_i$ snowman.
- Pirate throws$_m$/hears$_i$ guitar.
- Cook stirs$_m$/smells$_i$ soup.
- Leprechaun hits$_m$/feels$_i$ book.
- Witch climbs$_m$/builds$_i$ house.

The videos are available on http://www.phil.uu.nl/∼mariekes/ic-items/.[2]

[1] In the first interpretation a witch painting an existing table is meant; in the second a witch painting a table on a canvas.

[2] At least, they are at the time that this dissertation is published; if this url is no longer available, please contact the author.

Appendix B: Stimuli Chapter 6

Pilot study

Temporal adverbs:

- tomorrow
- yesterday
- at six o'clock
- at night

Sentences combined with temporal adverbs:

- The vase breaks.
- The boy rides a bicycle.
- The box falls over.
- The king takes a shower.
- The chicken moves in circles.

Temporal information experiment

The experiment used the images of the production experiment in chapter 5. These images were combined with an image of a clock, such as in the following example.

The following is a list of descriptions of all items.

- A gnome builds a building at seven o'clock.
- A pirate thinks of a shopping cart at six o'clock.
- A cook throws a saxophone at eight o'clock.
- A pirate swings a guitar at nine o'clock.
- A gnome eats pizza at three o'clock.
- A princess hangs a scarf at nine o'clock.
- A cook cuts a sock at five o'clock.
- A cook thinks of a sock at six o'clock.
- A pirate searches for a ball at one o'clock.
- A gnome sees a building at one o'clock.
- A cook hears a saxophone at two o'clock.
- A princess drops a vase at four o'clock.
- A princess knits a scarf at six o'clock.
- A pirate thinks of a guitar at ten o'clock.
- A princess dreams of a car at ten o'clock.
- A gnome pushes a cart at two o'clock.

Bibliography

Anderson, S. R. (2004). *Dr. Doolittle's Delusion*. Yale University Press, New Haven, CT.

Arbib, M. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, **28**, 105–167.

Arbib, M. A. and Bickerton, D., editors (2010). *The emergence of protolanguage: holophrasis vs compositionality*. John Benjamins.

Atkin, A. (2010). Peirceś theory of signs. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2010 edition.

Avramides, A. (1997). Intention and convention. In B. Hale and C. Wright, editors, *A Companion to the Philosophy of Language*. Blackwell.

Aydede, M. (2010). The language of thought hypothesis. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2010 edition.

Bach, K. (1997). The semantics-pragmatics distinction: What it is and why it matters. *Linguistische Berichte*, **8**, 33–50.

Balter, M. (2010). Animal communication helps reveal roots of language. *Science*, **328**, 969–971.

Benazzo, S. (2009). The emergence of temporality. In R. Botha and H. de Swart, editors, *Language evolution: the view from restricted linguistic systems*, LOT occasional series. LOT.

Berwick, R. (2011). Syntax facit saltum redux: Biolinguistics and the leap to syntax. In A. M. D. Sciullo and C. Boeckx, editors, *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford University Press.

Bickerton, D. (1981). *Roots of language*. Karoma, Ann Arbor, MI.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press, Chicago.

Bickerton, D. (1998). *Language and Human Behavior*. University of Washington Press, Seattle.

Bickerton, D. (2008). Ancestors of meaning. *Language and Communication*, **28**, 282–290.

Bidese, E., Padovan, A., and Tomaselli, A. (2012). Against protolanguage. In L. McCrohon, T. Fujimura, K. Fujita, R. Martin, K. Okanoya, R. Suzuki, and N. Yusa, editors, *Five approaches to Language Evolution: proceedings of the Workshops of the 9th International Conference on the Evolution of Language*.

Bolhuis, J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, **11**, 747–759.

Botha, R. (2003). *Unraveling the Evolution of Language*. Elsevier, New York.

Botha, R. (2006a). On the Windows Approach to language evolution. *Language and Communication*, **26**, 129–143.

Botha, R. (2006b). Pidgin languages as a putative window on language evolution. *Language and Communication*, **26**, 1–14.

Botha, R. (2007). On homesign systems as a potential window on language evolution. *Language and Communication*, **27**(1).

Botha, R. (2009a). On musilanguage/'hmmmmm' as an evolutionary precursor to language. *Language and Communication*, **29**, 61–76.

Botha, R. (2009b). What are windows on language evolution? In R. Botha and H. de Swart, editors, *Language evolution: the view from restricted linguistic systems*, LOT occasional series. LOT.

Botha, R. and de Swart, H., editors (2009). *Language evolution: the view from restricted linguistic systems*. LOT occasional series. LOT.

Calvin, W. and Bickerton, D. (2000). *Lingua ex machina: Reconciling Darwin and Chomsky with the Human Brain*. A Bradford Book.

Carstairs-McCarthy, A. (2005). The evolutionary origin of morphology. In M. Tallerman, editor, *Language origins: perspectives on evolution*, pages 166–184. Oxford University Press.

Carstairs-McCarthy, A. (2010). *The evolution of morphology*. Oxford University Press.

Casati, R. and Varzi, A. (2010). Events. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2010 edition.

Chomsky, N. (1988). *Language and Problems of Knowledge*. MIT press, Cambridge, MA.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Corballis, M. C. (2002). *From hand to mouth*. Princeton University Press.

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

Crist, E. (2004). Can an insect speak? The case of the honeybee dance language. *Social Studies of Science*, **34**(1), 7–43.

de Boer, B. (2001). *The origins of vowel systems*. Oxford University Press, Oxford.

de Swart, H. (1999). Position and meaning: time adverbials in context. In P. Bosch and R. van der Sandt, editors, *Focus: linguistic, cognitive and computational perspectives*, pages 336–361. Cambridge University Press.

de Swart, H. (2009). Negation in early L2: a window on language genesis. In R. Botha and H. de Swart, editors, *Language evolution: the view from restricted linguistic systems*, LOT occasional series, chapter 3. LOT.

de Swart, H. and de Hoop, H. (2000). Topic and focus. In *The First Glot International State-Of-The-Article Book*. Walter de Gruyter.

Dowty, D. R. (1989). On the semantic content of the notion of 'thematic role'. In G. Chierchia, B. H. Partee, and R. Turner, editors, *Properties, types and meaning*. Kluwer Academic Publishers.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, **67**(3), 547–619.

Dryer, M. S. (2011). Order of subject, object and verb. In M. S. Dryer and M. Haspelmath, editors, *The world atlas of language structures online*, Max Planck Digital Library.

Everaert, M. and Huybregts, R. (2012). The design principles of natural language. In J. J. Bolhuis and M. Everaert, editors, *Birdsong, Speech & Language. Converging Mechanisms*. MIT Press, Cambridge, MA.

Fay, N., Garrod, S., Roberts, L., and Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, **34**, 351–386.

Fitch, W. T. (2005). The evolution of language: a comparative review. *Biology and Psychology*, **20**, 193–200.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T. and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, **303**, 377–380.

Forbes, G. (2010). Intensional transitive verbs. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2010 edition.

Gamut, L. (1991). *Logic, Language and Meaning, Volume 2.* The University of Chicago Press.

Gärdenfors, P. (1993). The emergence of meaning. *Linguistics and Philosophy*, **16**, 285–309.

Gärdenfors, P. (1995). Language and the evolution of cognition. In *Lund University Cognitive Studies 41*.

Gärdenfors, P. (1996a). Cued and detached representations in animal cognition. *Behavioral Processes*, **36**, 263–273.

Gärdenfors, P. (1996b). Language and the evolution of cognition. In V. Rialle and D. Fisette, editors, *Penser l'esprit: Des sciences de la cognition à une philosophie cognitive*, pages 151–172. Presses Universitaires de Grenoble.

Gärdenfors, P. (1999). Some tenets of cognitive semantics. In J. Allwood and P. Gärdenfors, editors, *Cognitive Semantics: meaning and cognition*. John Benjamins.

Gärdenfors, P. (2004a). Cooperation and the evolution of symbolic communication. In D. K. Oller and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*. MIT Press.

Gärdenfors, P. (2004b). *How Homo Became Sapiens: On the evolution of thinking*. Oxford University Press.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, **31**, 961–987.

Gell-Mann, M. and Ruhlen, M. (2011). The origin and evolution of word order. *PNAS*, **108**(42).

Gentner, D. and Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman and S. Levinson, editors, *Language Acquisition and Conceptual Development*, pages 215–256. Cambridge University Press, New York.

Gentner, T., Fenn, K., Margoliash, D., and Nusbaum, H. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, **440**, 1204–1207.

Givon, T. (1997). On the co-evolution of language, mind and brain. *Evolution of communication*, **2**(1), 45–116.

Goldin-Meadow, S. (2003). Resilience of language. Psychology Press, New York.

Goldin-Meadow, S., So, W. C., Özyürek, A., and Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, **105**(27), 9163–9168.

Goldin-Meadow, S., Özyürek, A., Sancar, B., and Mylander, C. (2009). Making language around the globe: a cross-linguistic study of homesign in the United States, China, and Turkey. In J. Guo, E. Lieven, N. Budwig, and S. Ervin-Tripp, editors, *Crosslinguistic approaches to the psychology of language: research in the tradition of Dan Isaac Slobin*. Taylor and Francis, New York.

Gomez, F. and Miikkulainen, R. (1997). Incremental evolution of complex general behavior. *Adaptive behavior*, **5**, 317–342.

Gould, J. and Gould, C. (1995). *The Honey Bee*. Scientific American Library, New York.

Grandy, R. E. and Warner, R. (2009). Paul Grice. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2009 edition.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg, editor, *Universals of language*, pages 73–113. MIT press.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, San Diego, CA.

Grice, P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, **68**, 147–177.

Grice, P. (1982). Meaning revisited. In N. Smith, editor, *Mutual Knowledge*. Academic Press, New York.

Groenendijk, J. and Stokhof, M. (2005). Why compositionality? In G. Carlson and J. Pelletier, editors, *Reference and Quantification: The Partee Effect*. CSLI Press, Stanford.

Harvey, P. and Pagel, M. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, **298**, 1569–1579.

Heine, B. and Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, **203**, 88–96.

Holm, J. (1988). *Pidgins and creoles. Volume 1, theory and structure*. Cambridge University Press.

Hurford, J. R. (2007). *The Origins of Meaning*, volume I of *Language in the light of evolution*. Oxford University Press.

Hurford, J. R. (2011). *The Origins of Grammar*, volume II of *Language in the light of evolution*. Oxford University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, **3**(7), 272–279.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.

Johansson, S. (2005). *Origins of Language: constraints on hypotheses*, volume 5 of *Converging Evidence in Language and Communication Research*. John Benjamins, Amsterdam/Philadelphia.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press.

Kirby, S. (2007). The evolution of language. In R. Dunbar and L. Barrett, editors, *Oxford Handbook of Evolutionary Psychology*, pages 669–681. Oxford University Press.

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *PNAS*, **104**(12), 5241–5245.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory. *PNAS*, **105**(31).

Klein, W. (2001). Elementary forms of linguistic organization. In J. Trabant and S. Ward, editors, *New Essays on the Origin of Language*, pages 81–102. de Gruyter.

Klein, W. and Perdue, C. (1997). The Basic Variety (or: couldn't natural languages be much simpler?). *Second Language Research*, **13**, 301–347.

Kosslyn, S. (1994). *Image and brain*. MIT Press, Cambridge, MA.

Langus, A. and Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, **60**(4).

Lefebvre, C. (2009). The contribution of relexification, grammaticalisation, and reanalysis to creole genesis and development. In J. Clements and S. Gooden, editors, *Language Change in Contact Languages*, pages 277–311. John Benjamins.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.

Lewis, D. (1970). General semantics. *Synthese*, **22**, 18–67.

Lewis, D. (1975). Languages and language. In K. Gunderson, editor, *Minnesota Studies in the Philosophy of Science*, volume VII, pages 3–35. University of Minnesota Press.

McWhorter, J. (1997). *Towards a New Model of Creole Genesis*. Peter Lang, New York.

Meir, I., Lifshitz, A., Ilkbasaran, D., and Padden, C. (2010). The interaction of animacy and word order in human languages. In A. Smith, M. Schouwstra, B. de Boer, and K. Smith, editors, *The Evolution of Language*, pages 455–456, Singapore. World Scientific.

Millikan, R. G. (2004a). On reading signs: Some differences between us and the others. In K. Oller and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 15–30. MIT Press, Cambridge, MA.

Millikan, R. G. (2004b). *Varieties of Meaning*. MIT press.

Millikan, R. G. (2005). *Language: a Biological Model*. Oxford University Press.

Moltmann, F. (2008). Intensional verbs and their intentional objects. *Natural Language Semantics*, **16**(3), 239–270.

Morford, J. and Goldin-Meadow, S. (1997). From here and now to there and then: The development of displaced reference in homesign and English. *Child development*, **68**(3), 420–435.

Mufwene, S. S. (2007). What do creoles and pidgins tell us about the evolution of language? In B. Laks, S. Cleuziou, J.-P. Demoule, and P. Encreve, editors, *The origin and evolution of languages: approaches, models, paradigms*. Equinox, London.

Navarro, A. and Barton, N. H. (2003). Chromosomal speciation and molecular divergence — accelerated evolution in rearranged chromosomes. *Science*, **300**(5617), 321–324.

Newmeyer, F. J. (2000). On the reconstruction of 'proto world' word order. In C. Knight, J. R. Hurford, and M. Studdert-Kennedy, editors, *The evolutionary emergence of language*. Cambridge University Press, Cambridge.

Ohms, V. R., van Heijningen, C. A., Gill, A., Beckers, G. J., and ten Cate, C. (2010). Speaker-independent perception of human speech by zebra finches. In A. D. Smith, M. Schouwstra, B. de Boer, and K. Smith, editors, *The Evolution of Language*. World Scientific.

Oller, D. K. and Griebel, U. (2004). Theoretical and methodological tools for comparison and evolutionary modeling of communication systems. In D. K. Oller and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*. MIT Press.

Padden, C., Meir, I., Sandler, W., and Aronoff, M. (2010). *Against all expectations: Encoding subjects and objects in a new language*. MIT Press, Cambridge, MA.

Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*. MIT Press, Cambridge, MA.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, **31**, 109–178.

Perdue, C. (1993). *Adult Language Acquisition: Cross-Linguistic Perspectives*. Cambridge University Press.

Piaget, J. (1954). *The construction of reality in the child*. Routledge.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, **95**(2), 201–236.

Price, C. (2007). Language: A biological model (review). *Mind*, **116**(463), 766–769.

Reichenbach, H. (1947). *Elements of symbolic logic*, chapter 51, pages 287–298. the Macmillan Company.

Rescorla, M. (2011). Convention. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2011 edition.

Roberge, P. T. (2009). The creation of pidgins as a possible window on language evolution. In R. Botha and H. de Swart, editors, *Language evolution: the view from restricted linguistic systems*, LOT occasional series. LOT.

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *PNAS*, **102**(7).

Savage-Rumbaugh, E., Murphy, J., Sevcik, R., Brakke, K., Williams, S., Rumbaugh, D., and Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, **58**, 1–252.

Schel, A., Zuberbühler, K., and Slocombe, K. (2010). Why do wild chimpanzees produce food-associated calls: a case of vocal grooming? In A. D. Smith, M. Schouwstra, B. de Boer, and K. Smith, editors, *The Evolution of Language*. World Scientific.

Schouwstra, M. (2008). On there and then: from object permanence to displaced reference. In A. D. M. Smith, K. Smith, and R. F. i Cancho, editors, *The Evolution of Language*. World Scientific.

Schouwstra, M. (2010). Language evolution: the view from second language learners. In A. Smith, M. Schouwstra, B. de Boer, and K. Smith, editors, *The Evolution of Language*, Singapore. World Scientific.

Schouwstra, M., van Leeuwen, A., Smit, M., Marien, N., and de Swart, H. (2011). Semantic structure in improvised communication. In L. Carlson, C. Holscher, and T. Shipley, editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Austin, TX. Cognitive Science Society.

Scott-Phillips, T. C. and Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, **14**(9), 411–417.

Senghas, A., Coppola, M., Newport, E., and Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: the emergence of grammatical devices. In E. Hughes, M. Hughes, and A. Greenhill, editors, *Proceedings of the Boston University conference on language development*, pages 550–561, Boston. Cascadilla Press.

Senghas, A., Özyürek, A., and Kita, S. (2002). Encoding motion events in an emerging sign language: From Nicaraguan gestures to Nicaraguan signs. In A. Baker, B. van den Bogaerde, and O. Crasborn, editors, *Cross-linguistic perspectives in sign language research. Selected papers from TISLR 2000.* Signum Press.

Senghas, A., Kita, S., and Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, **305**(5691), 1779–1782.

Sjölander, S. (1997). On the evolution of reality — some biological prerequisites and evolutionary stages. *Journal of Theoretical Biology*, **187**, 595–600.

Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz and S. Levinson, editors, *Rethinking linguistic relativity.* Cambridge University Press.

Smith, K. (2006). The protolanguage debate: bridging the gap? In A. Cangelosi, A. Smith, and K. Smith, editors, *The Evolution of Language*. World Scientific.

Smith, K. (2010). Is a holistic protolanguage a plausible precursor to language? A test case for modern evolutionary linguistics. In M. A. Arbib and D. Bickerton, editors, *The emergence of protolanguage: holophrasis vs compositionality*. John Benjamins.

Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in complex systems*, **6**(4), 537–558.

Speaks, J. (2010). Theories of meaning. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2010 edition.

Starren, M. (2001). *The Second Time. The Acquisition of Temporality in Dutch and French as a second language*. Ph.D. thesis, Netherlands Graduate School of Linguistics.

Szabó, Z. G. (2008). Compositionality. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, **117**(3), 579–604.

Tinbergen, N. (1964). The evolution of signalling devices. In W. Etkin, editor, *Social Behavior and Organization among Vertebrates*. University of Chicago Press, Chicago.

Tomasello, M. (2001). *The Cultural Origins of Human Cognition*. Harvard University Press.

van Leeuwen, A. (2010). *Gesturing orders and their interpretation*. Master's thesis, Utrecht University.

Vendler, Z. (1957). Verbs and times. *Philosophical review*, **66**(143–160).

Verhoef, T., de Boer, B., and Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In T. C. Scott-Phillips, M. Tamariz, E. Cartmill, and J. Hurford, editors, *The evolution of language: Proceedings of the 9th international conference (evolang9)*.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, **18**.

Wynne, C. D. (2004). *Do animals think?* Princeton University Press.

Zucchi, S. (1999). Incomplete events, intensionality and imperfective aspect. *Natural Language Semantics*, **7**, 179–215.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, Cambridge, MA. MIT Press.

Zuidema, W. (2005). *The Major Transitions in the Evolution of Language*. Ph.D. thesis, Theoretical and Applied Linguistics, University of Edinburgh, UK.

Samenvatting in het Nederlands

Semantische prototaal en beperkte taalsystemen

Mijn proefschrift biedt een perspectief op taalevolutie waarin betekenis centraal staat. Betekenis krijgt in het debat over het ontstaan van taal weinig aandacht, omdat men zich doorgaans vooral op de (complexe) syntactische structuur van taal concentreert. Ik laat juist zien dat mijn aanpak vanuit betekenis op een zinnige manier aan dat debat kan bijdragen. Met betekenis van taal bedoel ik niet alleen de betekenis van individuele woorden, maar vooral de betekenis van combinaties van woorden: taal heeft de bijzondere eigenschap dat complexe betekenissen door sprekers van een taal gebouwd kunnen worden door woorden met eenvoudige betekenissen samen te voegen. Mijn proefschrift onderzoekt de hypothese dat er, voorafgaand aan syntactisch complexe taal, een stadium is geweest waarin taal werd gestuurd door semantische principes. Als dit stadium er inderdaad geweest is, zou dat een brug slaan tussen primitieve communicatie en moderne, complexe taal.

Ik geef deze hypothese vorm door een claim te maken over prototaal, en te zeggen dat prototaal semantisch georganiseerd was. Prototaal is een hypothetisch tussenstadium in het ontstaan van taal en in de literatuur zijn veel voorstellen gedaan over het karakter van prototaal (bijvoorbeeld: holistische, lexicale en muzikale prototaal). Hoofdstuk 1 van dit proefschrift geeft een overzicht van deze invullingen en concludeert dat de notie van prototaal een nuttig conceptueel hulpmiddel is om hypothesen op te stellen over de evolutionaire geschiedenis van taal. Ik concludeer bovendien dat de verschillende vormen van prototaal mogelijkerwijs naast elkaar hebben bestaan, maar geef daarbij aan dat ik mij verder zal concentreren op *semantische prototaal*.

Semantische prototaal is terug te voeren op het idee van Jackendoff (2002), dat in een bepaald stadium in het ontstaan van taal semantische en pragmatische principes als AgentFirst en FocusLast een rol speelden in het structureren van talige uitingen. In zijn boek 'Foundations of Language' geeft Jackendoff al enkele voorbeelden van bewijs voor het bestaan van die principes, maar

het doel van mijn proefschrift is om een grondig onderzoek te starten naar de levensvatbaarheid van de semantische prototaal-hypothese.

Hoofdstuk 2 laat verschillende manieren zien om empirisch bewijs te verzamelen over het ontstaan van taal. Dit bewijs is noodzakelijkerwijs *indirect*, omdat we nu eenmaal geen opnamen hebben van het taalgebruik van onze evolutionaire voorouders. De methode die centraal staat in dit proefschrift is het kijken naar *beperkte taalsystemen*, om daaruit conclusies te trekken over evolutionair vroege taal (prototaal). Ik geef een overzicht van observaties van pidgintalen (taalsystemen ontwikkeld in situaties waarin verschillende groepen geen gemeenschappelijke taal hebben en toch willen communiceren), homesign (gebaarsystemen ontwikkeld door dove kinderen die opgroeien in gezinnen waarin geen gebarentaal wordt gesproken), de Basic Variety (een stadium in het leerproces van volwassen tweede taalleerders) en spontaan ontstane gebarentalen in dovengemeenschappen. De verschillende situaties waarin deze observaties zijn gedaan hebben gemeen dat taalgebruikers geen gebruik kunnen maken van hun moedertaal; ofwel omdat hun toehoorders die niet zullen begrijpen (zoals dat bij tweede taalleerders het geval is), ofwel omdat ze niet in staat zijn een moedertaal te leren (zoals bij homesigners). Uitingen in deze systemen zijn vaak kort, en bevatten doorgaans weinig grammaticale elementen, zoals functiewoorden, naamvallen of werkwoordsvervoegingen. Toch zijn ze redelijk succesvol voor communicatie.

Ik laat zien dat de systemen overeenkomsten vertonen, en dat Jackendoffs principes AgentFirst en FocusLast inderdaad terug te zien zijn. De algemene indruk is dus dat als mensen niet hun eigen taal kunnen gebruiken, ze hun uitingen wel degelijk structuur geven. De organisatorische principes die ze gebruiken, lijken inderdaad semantisch van aard.

Er is echter geen goede manier om de verschillende systemen grondig met elkaar te vergelijken, en bij veel van de empirische data die voorhanden is zijn grote meningsverschillen over de interpretatie ervan. Mijn voornaamste conclusie is dan ook dat, om semantische prototaal een stevige empirische basis te geven, er meer empirisch materiaal nodig is. Het verzamelen van data uit beperkte taalsystemen is alleen erg bewerkelijk. Bovendien zou het mooi zijn als we data zouden kunnen verzamelen in een meer gecontroleerde omgeving, zodat we heel specifieke hypotheses kunnen testen. Aan het slot van hoofdstuk 2 kondig ik een methode aan die dit biedt: experimenten met geïmproviseerde communicatie in het lab. Op deze methode kom ik in het laatste deel van mijn proefschrift terug; het middelste gedeelte concentreert zich op een gedetailleerde analyse van het begrip betekenis. Zo'n analyse geeft ons de mogelijkheid om in preciezere termen te praten over de empirische data en de conclusies over de evolutie van taal die daaruit getrokken worden.

Betekenis en evolutie

Hoofdstuk 3 buigt zich over definities van betekenis. Het antwoord dat men zal krijgen op een vraag als 'wat is betekenis?' hangt af van degene die dat antwoord geeft. Ik kijk naar definities van betekenis in drie verschillende contexten: die van de filosofen, die van denkers over evolutie en die van taalkundigen.

In de filosofie is betekenis een veel bestudeerd fenomeen. Uit een overzicht van verschillende definities destilleer ik drie centrale intuïties over de essentie van betekenis:

referentie De dingen in de wereld waarnaar woorden en zinnen verwijzen zijn belangrijk als we iets willen weten over hun betekenis.

intentie en overtuiging Als een spreker een uiting (zin) formuleert, heeft hij doorgaans de intentie om zijn toehoorders tot een bepaalde overtuiging te laten komen.

conventie Onze talige uitingen hebben de betekenis die ze hebben omdat ze zo gebruikt zijn door andere sprekers in gelijkwaardige situaties.

Ik gebruik deze intuïties niet als competitieve noties, maar als verschillende manieren om naar betekenis te kijken, waarbij ik betekenis zie als een fenomeen met verschillende facetten. Ze zullen in de latere hoofdstukken van mijn proefschrift dan ook terugkomen.

Als we naar betekenis kijken vanuit een evolutionair perspectief, zien we dat er andere eisen worden gesteld aan definities ervan. We kunnen niet zomaar aannemen dat betekenissen abstracte objecten zijn, maar we moeten een verklaring geven voor de evolutionaire geschiedenis ervan. Ik bekijk twee mogelijke scenario's voor het ontstaan van betekenis in menselijke taal: de eerste werd voorgesteld door Peter Gärdenfors; de tweede door Ruth Millikan.

Gärdenfors gaat er, kort gezegd, vanuit dat cognitie is ontwikkeld voor taal: private (persoonlijke) mentale representaties worden gezien als de basis voor geavanceerde communicatie. Een gevolg van deze aanname is dat mensapen (onze naaste verwanten in het dierenrijk) volgens Gärdenfors heel goed ver ontwikkelde cognitieve vermogens zouden kunnen hebben, die ze echter niet voor communicatie gebruiken. Andere auteurs die schrijven over de evolutionaire oorsprong van betekenis lijken (impliciet) uit te gaan van het scenario zoals Gärdenfors het voorstelt, en er wordt veel vergelijkend onderzoek uitgevoerd naar de cognitieve vermogens van mensen en van andere dieren. Er is echter ook een manier om te kijken naar het ontstaan van betekenis en taal, zonder cognitie zo centraal te stellen. Die komt naar voren in het werk van Ruth Millikan.

Millikan concentreert zich in haar werk op conventies en legt daarbij de nadruk op taal als publiek fenomeen. Ze ontkent dat er een een-op-een-relatie is tussen de betekenis van een zin en de mentale representatie van degene die zo'n zin uitspreekt. In plaats daarvan benadrukt ze de conventionele kant van taal (zie het kader hierboven voor een definitie), en geeft ze een biologische fundering voor conventies in menselijk gedrag.

Cognitie en communicatie 'in het wild' en in het lab

Gewapend met de principes uit hoofdstuk 3 concentreer ik me in hoofdstuk 4 weer op semantische prototaal. Observaties uit beperkte taalsystemen kunnen gebruikt worden als bewijs voor het bestaan van zo'n stadium, maar om de problemen uit hoofdstuk 2 het hoofd te bieden, wend ik me tot een nieuwe experimentele methode: geïmproviseerde communicatie. Volgens deze methode wordt aan proefpersonen gevraagd om te communiceren over eenvoudige afbeeldingen, zonder te spreken en door middel van gebaren (uitbeelden). Eerdere toepassingen van deze methode lieten zien dat proefpersonen bij het uitbeelden een volgorde gebruikten die afweek van de dominante woordvolgorde van hun moedertaal. Bovendien gebruikten sprekers van verschillende moedertalen altijd dezelfde volgorde: *Actor-Patient-Act* (Goldin-Meadow *et al.*, 2008). Deze volgorde is consistent met de woordvolgorde Subject-Lijdend voorwerp-Object, en dit wil zeggen dat een gebeurtenis waarin een man een bal gooit, steevast wordt gecommuniceerd als 'man-bal-gooien'.

Voordat ik laat zien dat semantische principes inderdaad een rol spelen in geïmproviseerde communicatie, ga ik in op de conclusies over taalevolutie die we op basis van de experimenten zouden kunnen trekken.

Ik vergelijk de situatie waarin proefpersonen in het experiment verkeren met de situatie van onze evolutionaire voorouders. Zowel de proefpersonen in het experiment als onze voorouders bevinden zich in een situatie waarin er nog geen volledig ontwikkeld systeem van linguïstische conventies is. In beide situaties zullen sprekers (in het geval van het experiment is het beter om te spreken over gebaarders, maar ik zal ook hier de term sprekers gebruiken) een bepaalde mentale representatie hebben, die ze aan iemand duidelijk willen maken. Het verschil tussen een mentale representatie en een zin (of een serie gebaren) is de vorm. Een mentale representatie kan best de vorm van een afbeelding hebben, maar als je iets zegt of uitbeeldt, moet je noodgedwongen een aaneenschakeling van symbolen maken (een *string*). Bij het omzetten van een mentale representatie naar een communicatieve uiting spelen cognitieve voorkeuren een rol, en deze cognitieve voorkeuren vallen, zo beweer ik, nagenoeg samen met de semantische principes die eerder genoemd werden voor semantische prototaal.

Het feit dat proefpersonen in het experiment geen gebruik maken van de grammaticale regels van hun moedertaal wijst erop dat ze blijkbaar gebruik

maken van een fundamenteler systeem, een systeem dat onderliggend is aan taal.

Geïmproviseerde communicatie

In de laatste twee hoofdstukken van mijn proefschrift, hoofdstuk 5 en 6, doe ik verslag van experimenteel onderzoek. In hoofdstuk 5 begin ik bij de eerder gevonden resultaten waarin er een vaste volgorde werd gevonden in de geïmproviseerde communicatie over eenvoudige gebeurtenissen: *Actor-Patient-Act* (Goldin-Meadow *et al.*, 2008). Ik laat zien dat er naast de gebeurtenissen die werden gebruikt in dit experiment (extensionele gebeurtenissen), nog andere gebeurtenissen zijn, die andere semantische eigenschappen hebben: intensionele gebeurtenissen. Intensionele gebeurtenissen zijn abstracter dan extensionele gebeurtenissen, omdat ze gaan over dingen die niet per se aanwezig zijn, of die niet specifiek zijn. Een paar voorbeelden: 'Een man gooit een bal' is een extensionele gebeurtenis, maar 'Een man zoekt een bal' is een intensionele gebeurtenis. Dit is zo, omdat het mogelijk is dat de man niet naar *een bepaalde bal* op zoek is.

In een eerste experiment laat ik zien dat proefpersonen verschillende gebarenvolgordes gebruiken voor de twee soorten gebeurtenissen: Actor-Patient-Act voor extensionele gebeurtenissen (dat is gelijk aan het eerder uitgevoerde onderzoek) en Actor-Act-Patient voor intensionele gebeurtenissen. Met andere woorden, als een proefpersoon een situatie moet uitbeelden waarin een man een bal zoekt, zal hij dat steevast doen in deze volgorde: 'man-zoeken-bal'. Ik laat daarmee zien dat verschillende semantische eigenschappen van datgene waarover gecommuniceerd wordt, leiden tot verschillende woordvolgordes.

Om te laten zien dat deze volgordes een communicatieve functie hebben, presenteer ik een tweede experiment, waarin proefpersonen gevraagd werd video's met series gebaren te interpreteren. De acties in elke video waren ambigu, zodat elke *string* van gebaren twee dingen zou kunnen betekenen; zie de afbeeldingen op pagina 145. De video's werden in twee volgordes gepresenteerd (Actor-Patient-Act of Actor-Act-Patient) en proefpersonen kregen na elke *string* de keuze uit twee interpretaties. Het onderzoek laat zien dat de volgorde van de gebaren inderdaad invloed heeft op de interpretatie van de ambigue gebaren. Dit wijst erop dat het veranderen van de gebarenvolgorde voor extensionele en intensionele gebeurtenissen een communicatieve functie heeft.

Hoofdstuk 6 onderzoekt wat er gebeurt als proefpersonen complexe informatie moeten overbrengen door middel van geïmproviseerde communicatie. Deze complexe informatie ontstaat bijvoorbeeld als er aan de informatie over een gebeurtenis informatie over het tijdstip van de gebeurtenis wordt toegevoegd. In bestaande talen gebruiken we veelal werkwoordsvervoegingen om aan te duiden of een bepaalde gebeurtenis in het verleden, heden de toekomst plaatsvindt.

Van beperkte taalsystemen weten we dat werkwoordvervoeging er niet mogelijk is. In die systemen wordt echter wel gepraat over verleden en toekomst.

Sprekers gebruiken temporele adverbia aan het begin van de zin, om aan te geven dat gebeurtenissen niet nu plaatsvinden, maar op een tijdstip in het verleden of de toekomst (Benazzo, 2009).

Ik presenteer een experiment met geïmproviseerde communicatie, waarin proefpersonen over eenvoudige gebeurtenissen, plus de tijd waarop deze gebeurtenissen plaatsvonden, moesten communiceren. Het experiment laat zien dat mensen in het lab dezelfde methode toepassen als die was geobserveerd in beperkte taalsystemen.

Hoofdstuk 7 geeft een overzicht van de belangrijkste conclusies van dit proefschrift. De hypothese dat er, voorafgaand aan syntactisch complexe taal, een stadium is geweest waarin taal werd gestuurd door semantische principes wordt niet langer alleen ondersteund door empirische observaties uit beperkte taalsystemen, maar ook door resultaten van experimenten. Daarmee heb ik een bijdrage geleverd aan het debat over het ontstaan van taal en onmiskenbaar duidelijk gemaakt dat de rol van betekenis in het ontstaan van taal niet genegeerd kan worden. Om dat te bereiken heb ik een experimentele methode (geïmproviseerde communicatie) verder ontwikkeld en een brug geslagen tussen deze methode en de eerder genoemde data uit beperkte taalsystemen.

Curriculum Vitae

Marieke Schouwstra was born on April 2nd, 1981 in Harderwijk, the Netherlands. From 1999 to 2006 she studied Cognitive Artificial Intelligence in Utrecht, where she specialised in logic and language. After doing a research internship at the ILLC at the University of Amsterdam studying dynamic semantics, she obtained her master's degree with a thesis on possible worlds accounts of meaning and the problem of logical omniscience. In 2007 she started a PhD project at the Utrecht Institute of Linguistics OTS, focusing on the evolution of language and meaning. In 2009 she spent a semester as a visiting PhD student at the Language Evolution and Computation group at the University of Edinburgh. This dissertation is the result of her research as a PhD student. Marieke is currently employed as a lecturer at Utrecht University.